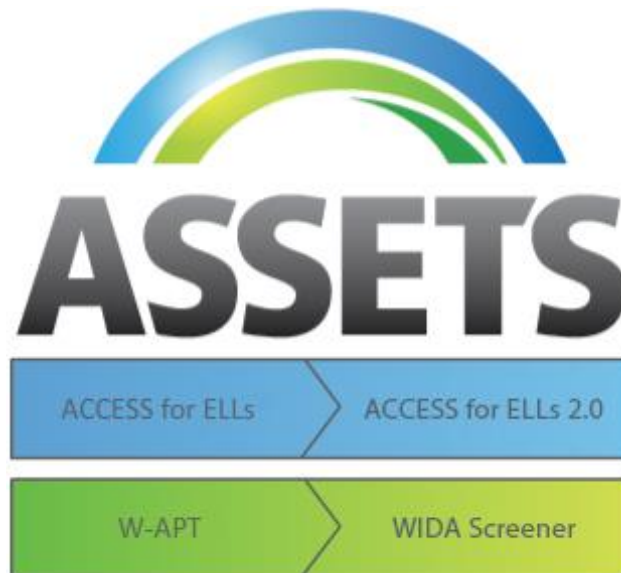


# The WIDA Consortium English Language Proficiency Assessment for Grades 1-12:

## Test and Item Design Plan for the Annual Summative and On-demand Screener

Last updated  
October 22, 2014

Prepared by:  
Center for Applied Linguistics



## Contents

1.0 Introduction .....	4
1.1 Scope of this document .....	4
1.1.1 Summative Assessment and Screener .....	4
2.0 Background .....	6
2.1 The ASSETS Project.....	6
2.2 The WIDA Consortium.....	6
2.3 Grant-funded Assessment System .....	8
2.4 Levels of documentation.....	10
2.5 ACCESS and ACCESS 2.0 .....	11
3.0 ACCESS 2.0 at a Glance .....	13
4.0 Test-level Concerns .....	14
4.1 Screener Functionality .....	14
4.2 Design and functionality .....	15
4.2.1 General principles .....	15
4.2.2 Design elements.....	16
4.2.3 Navigation .....	17
4.3 Accessibility and affect.....	18
4.3.1 Instructions and Item Components .....	18
4.3.2 Virtual test administrator.....	20
4.3.3 Variation across grade levels and language proficiency levels .....	21
4.4 Key academic language uses.....	21
4.5 Adaptivity .....	22
4.5.1 Order of subtest administration .....	22
4.6 Video Tutorial and Student Practice Module.....	23
4.7 Paper-based accommodation.....	23
4.8 Reporting.....	25
4.8.1 Scores reported.....	25
4.8.2 Score reports.....	26
5.0 Subtest Specifics: Listening .....	27
5.1 Composition of the Listening subtest .....	27
5.2 Adaptivity .....	28
5.3 Item presentation .....	29
5.4 Domain-specific design elements .....	30
5.5 Student response .....	30

5.6 Scoring.....	31
5.7 Listening screener .....	31
6.0 Subtest Specifics: Reading.....	32
6.1 Composition of the Reading subtest.....	32
6.2 Adaptivity .....	32
6.3 Item presentation .....	33
6.4 Domain-specific design elements .....	33
6.5 Student response .....	34
6.6 Scoring.....	34
6.7 Reading screener.....	34
7.0 Subtest Specifics: Writing .....	35
7.1 Composition of the Writing subtest.....	35
7.2 Tier placement .....	35
7.3 Task presentation.....	36
7.4 Domain-specific design elements .....	37
7.5 Student response .....	37
7.6 Scoring.....	38
7.7 Writing screener .....	38
7.7.1 Composition of the Writing screener.....	38
7.7.2 Scoring the Writing screener .....	39
8.0 Subtest Specifics: Speaking .....	40
8.1 Composition of the Speaking subtest .....	40
8.2 Tier placement .....	40
8.3 Task presentation and student response.....	41
8.4 Domain-specific design elements .....	42
8.5 Scoring.....	44
8.6 Speaking screener .....	44
8.6.1 Composition of the Speaking screener .....	44
8.6.2 Scoring the Speaking screener .....	45
9.0 Next Steps .....	45
Appendix: Understanding the ACCESS for ELLs Assessment .....	47
A1.1 Introduction .....	47
A1.2 Test structure.....	47
A1.3 Tiers .....	48
A1.4 Test administration times .....	50

## 1.0 Introduction

This document, the *WIDA Consortium English Language Proficiency Assessment for Grades 1-12: Test and Item Design Plan for the Annual Summative and On-demand Screener*, builds on *The WIDA Consortium English Language Proficiency Assessment Framework: Annual Summative and On-demand Screener* to provide additional detail about (a) the new technology-delivered annual summative assessment of academic English language development, ACCESS for ELLs 2.0 (hereafter referred to as ACCESS 2.0), and (b) the on-demand WIDA Screener, a test given to incoming students who may be designated as English language learners.

Like the *Framework*, the *Test and Item Design Plan* (TIDP) is a living document that can and will evolve in response to feedback from stakeholders from throughout the WIDA Consortium. Note that this document describes the **operational version** of the ACCESS 2.0 and WIDA Screener assessments; due to vendor specifications, some minor details of the new assessments described in this document may not be fully present in the **field test version** of the assessments.

### 1.1 Scope of this document

The TIDP is complementary to the Framework. The TIDP includes more detail about some topics than the Framework, and the Framework contains background information absent from the TIDP. To increase the TIDP's ability to function as a stand-alone document, however, some sections from the Framework appear more or less verbatim in the TIDP. Readers familiar with the Framework may feel free to skip such sections. Readers for whom more background on the current ACCESS for ELLs assessment would inform their reading of this document should consult the appendix *Understanding the ACCESS for ELLs Assessment*.

#### 1.1.1 Summative Assessment and Screener

This document describes the annual summative and screener components of the new assessment system (see section 2.2). Table 1 below describes the purposes and uses of each assessment. Differences between the assessments will be noted throughout this document as necessary.

**Table 1: Differences between ACCESS 2.0 and the ACCESS 2.0 Screener**

Aspect of test	ACCESS 2.0	ACCESS 2.0 Screener	Where to find more information
<b>Purpose and Uses</b>	<ul style="list-style-type: none"> <li>secure, annual measure that meets federal requirements for monitoring ELLs' progress toward English language proficiency</li> <li>can serve as one of multiple measures used to determine student preparedness to exit English language support services</li> </ul>	<ul style="list-style-type: none"> <li>on-demand, locally-scored measure that meets federal requirements as an instrument to help determine a student's ELL status in a timely fashion</li> <li>can serve as one of multiple measures to determine if a student is an English language learner qualified for English language support services</li> </ul>	Framework 3
<b>Composition</b>	<ul style="list-style-type: none"> <li>Listening: 6-8 folders of test questions (17 to 26 questions)</li> <li>Reading: 8-10 folders (23 to 32 questions)</li> <li>Writing: 3 tasks, one of them extended (Tier B/C only)</li> <li>Speaking: 3 folders (6 tasks); for Pre-A, 3 folders (3 tasks)</li> </ul>	<ul style="list-style-type: none"> <li>Listening: 4-6 folders (11 to 20 questions)</li> <li>Reading: 4-6 folders (11 to 20 questions)</li> <li>Writing: 1 extended task</li> <li>Speaking: 1-3 folders (up to 6 tasks)</li> </ul>	TIDP 5.1, 5.7, 6.1, 6.7, 7.1, 7.7, 8.1, 8.6
<b>Scoring</b>	<ul style="list-style-type: none"> <li>Speaking responses centrally scored using a six-level rubric</li> <li>keyboarded Writing responses or scanned handwritten responses centrally scored</li> </ul>	<ul style="list-style-type: none"> <li>Speaking responses scored locally by educators trained using the ACCESS 2.0 Screener Rater Training Program—Speaking</li> <li>keyboarded Writing responses or handwritten responses scored locally by educators trained using the ACCESS 2.0 Screener Rater Training Program—Writing</li> </ul>	TIDP 7.6, 7.7.2, 8.5, 8.6.2

<b>Administration</b>	<ul style="list-style-type: none"> <li>test administrators need not be trained to score student responses for any domain</li> <li>test administrators need not be trained to score student responses for any domain</li> </ul>	Framework 8
<b>Reporting</b>	<ul style="list-style-type: none"> <li>electronic data files and score reports will be produced in and distributed from a central location for distribution through state and district offices as appropriate</li> <li>Speaking and Writing scores are entered into an interface that combines them with the stored Listening and Reading scores, calculates the various composite scores, and generates electronic data files and score reports for local use</li> </ul>	TIDP 4.7

## 2.0 Background

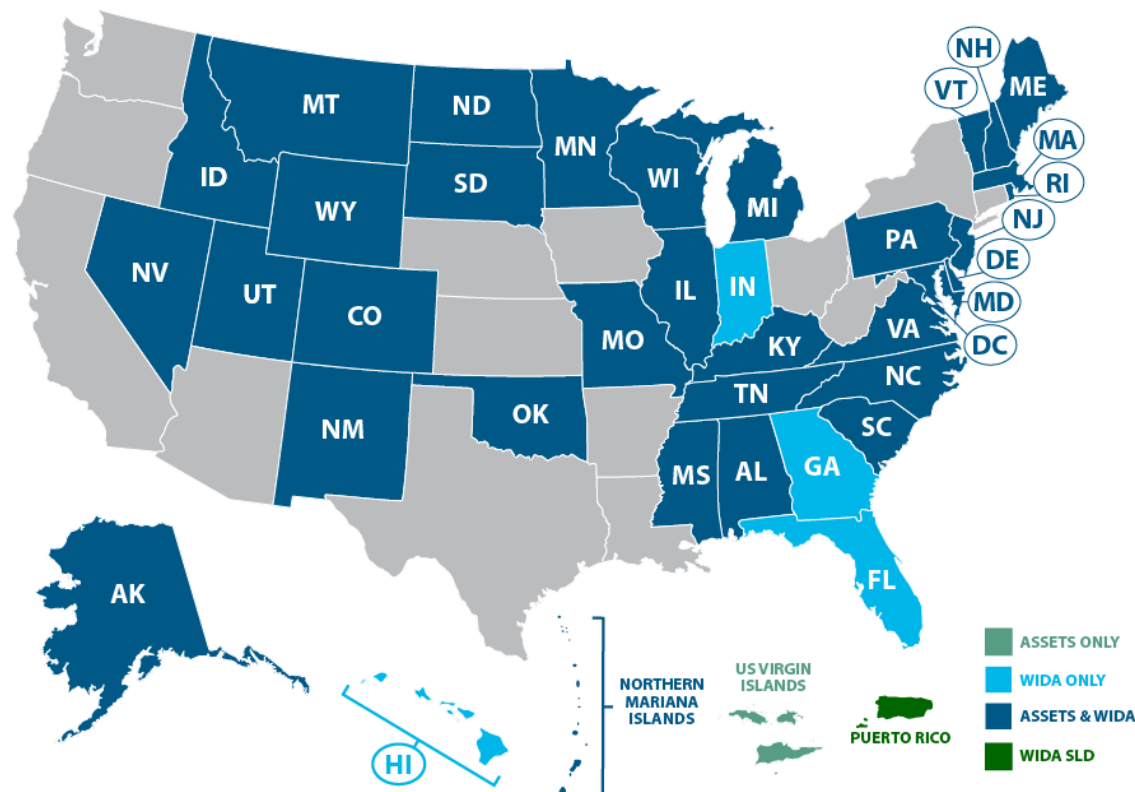
### 2.1 The ASSETS Project

The ASSETS Project is funded by a four year Enhanced Assessment Grant from the U.S. Department of Education. Between 2011 and 2015, WIDA, the Wisconsin Department of Public Instruction (Fiscal Agent), and project partners will develop a next generation English language proficiency (ELP) assessment system that will continue to be supported via states' participation in the WIDA Consortium post-2015.

### 2.2 The WIDA Consortium

The WIDA Consortium is housed within the Wisconsin Center for Education Research at the University of Wisconsin–Madison. It was established in 2003 with funding from a U.S. Department of Education Enhanced Assessment Grant, awarded to the state of Wisconsin. The WIDA Consortium currently includes 36 states and the District of Columbia. Figure 1 shows the states that are currently members of the WIDA Consortium. More information about their status is available on the WIDA website.

Figure 1: The WIDA Consortium



Since its beginning, the WIDA Consortium (hereafter WIDA) has created and expanded on comprehensive English language development (ELD) standards (2004, 2007, 2012) that represent the second language acquisition process. The five basic standards cover the language students need to comprehend and produce in five areas of academic English language: social and instructional language and the language of the content areas of language arts, mathematics, science, and social studies. Based on these standards, WIDA has developed the following assessments for use by WIDA states:

- a K–12 annual summative English language proficiency (ELP) test, Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs);
- an initial screener, the WIDA ACCESS Placement Test (W-APT); and
- an on-demand, “off-the-shelf” test of ELP known as WIDA MODEL that can be used for placement or for interim assessment both within WIDA states and also by non-WIDA educational programs.

In addition to its standards and assessments, WIDA pursues a research agenda on behalf of member states. WIDA research explores not only the validity of the assessments, but also areas of interest such as ELP growth rates, correlations between its ELP tests and state academic tests, and classroom implementation of the ELD standards. Concurrently, WIDA also provides extensive professional development opportunities and maintains a comprehensive website ([www.wida.us](http://www.wida.us)). The Central Office

of the WIDA Consortium serves as the management partner for the ASSETS Project in carrying out the grant activities.

## 2.3 Grant-funded Assessment System

The funding awarded for the ASSETS Project will allow the WIDA Consortium to develop a technology-enhanced assessment system to assess English learners' (ELs') acquisition of the academic English language needed for college and career success. By the end of the 2015-16 school year, the new assessments will be part of a comprehensive system that is (a) technology-based, incorporating several major technological enhancements; (b) anchored in the established WIDA ELD Standards, which correspond with college and career readiness standards including the Common Core State Standards, Next Generation Science Standards, and other state content standards; (c) informed by rigorous ongoing research; and (d) supported by comprehensive professional development and outreach, all of which will be developed within the framework of the multistate WIDA Consortium.

The comprehensive system will include:

- **English Language Development Standards:** The WIDA ELD Standards will be a key component of the new assessments. First published in 2004 and updated in 2007 and 2012, the standards capture an evolving understanding of (a) the needs of ELs and their educators and (b) the use of the standards as the foundation for instruction and assessment.
- **A common definition of English language learner:** A subcommittee of WIDA member states will be part of the national initiative to develop a common definition of what it means to be an English language learner (ELL). This common definition will be used by states to identify, classify, and reclassify ELLs.
- **English Language Proficiency assessments:** The ELP assessments will use technology to allow for (a) increased student engagement through a more dynamic testing experience; (b) built-in accommodations and accessibility features appropriate for a range of student needs; (c) simultaneous administration of multiple grades and tiers; and (d) increased ease for test administrators, who will no longer need to administer and score the Speaking test one-on-one. The new assessments to be developed are as follows:
  - **A computer-based summative test, ACCESS for ELLs 2.0:** ACCESS for ELLs 2.0—to be administered annually in grades 1–12<sup>1</sup> for accountability and program purposes—will cover (a) the language domains of listening, speaking, reading, and writing and (b) the five WIDA ELD Standards, encompassing social and instructional language and the language of language arts, mathematics, science, and social studies.
  - **A computer-based on-demand diagnostic (screener) test, the WIDA Screener:** The WIDA Screener will be used to determine eligibility for ELL services and program placement within those services. The test format will be derived from that of ACCESS for ELLs 2.0, but scoring not done automatically by computer will be done locally.

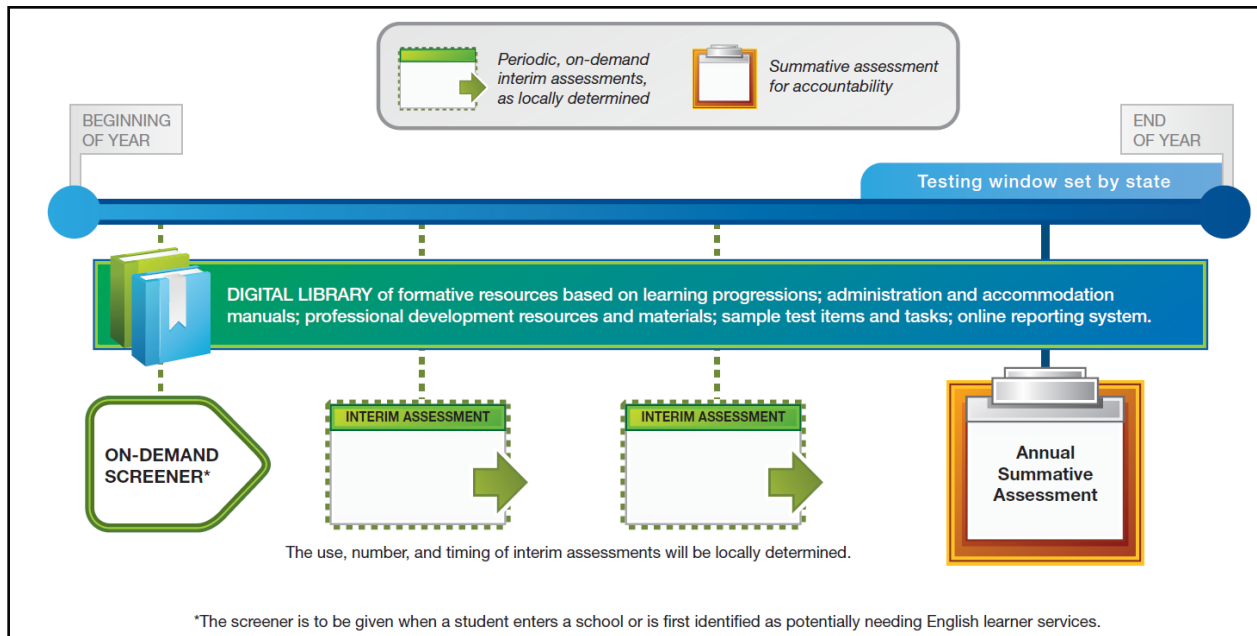
---

<sup>1</sup> Kindergarten is not included in the grant and will remain an interactive, paper-based kit for the near future.



- **Computer-based interim assessments:** A series of shorter, targeted interim assessments will be developed to enable schools to chart student progress in finer increments and with more precision than ACCESS for ELLs 2.0 allows, and to help guide instruction.
- **Formative assessment resources:** English language learning progressions will be researched and developed for both the academic and social English associated with academic success and career readiness. These progressions will provide a foundation for the development of formative assessment resources to help educators monitor student growth during instruction.
- **A training program for local scorers:** A scorer training program based on CAL's existing Multimedia Rater Training Program (MRTP) will provide intensive, on-demand training and practice in scoring speaking and writing performances on the screener and the interim assessments.
- **Professional development and outreach materials:** Materials will be developed for (a) professional development on how to administer the assessment system, including the appropriate and effective use of assessment results, and (b) outreach to stakeholders, including educators, state agencies, and policymakers.

Figure 2: The New Assessment System



([http://www.k12center.org/rsc/pdf/18901-ASETSConsortiumFlier\\_WEB.pdf](http://www.k12center.org/rsc/pdf/18901-ASETSConsortiumFlier_WEB.pdf))

It is important to note that the funding awarded for the ASSETS grant covers partial development of the new assessment system, with WIDA supporting additional resource development as well as the overall transition. All work is being done in close consultation with WIDA Consortium member states.

## 2.4 Levels of documentation

This document presents more detail about the annual summative and screener assessments outlined in broad strokes in *The WIDA Consortium English Language Proficiency Assessment Framework: Annual Summative and On-demand Screener*. Both the Framework and the TIDP are public documents geared primarily for representatives of State Education Agencies (SEAs) to use as consensus-building tools among Consortium members and as documentation and communication comprehensible to a general audience involved in the assessment of ELLs.

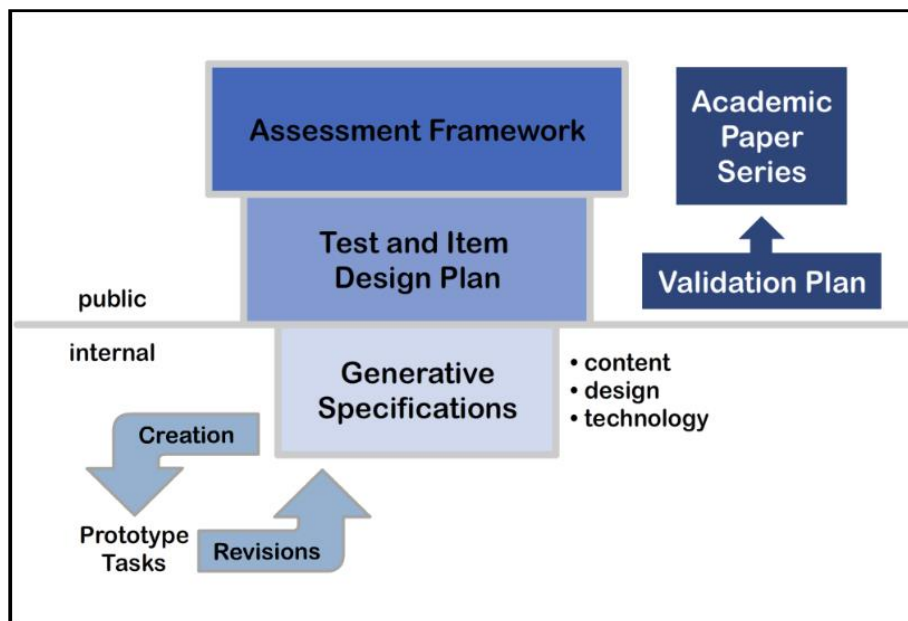
In addition to these consensus-building documents, the Center for Applied Linguistics (CAL), the main test development partner, and the WIDA Consortium plan to produce a series of academic research papers and reports. This academic paper series will explicate the theoretical foundations underlying the assessment system—the construct of Academic English, for example—document the findings of qualitative and quantitative research conducted during the various phases of its development, and provide technical psychometric analyses that constitute support for the use of the system and its components.

The selection of topics to be covered in the research papers and reports will be guided in part by a validation plan. The validity of a test is the extent to which evidence supports the interpretations of test scores associated with the proposed uses of the test, and the ACCESS 2.0 validation plan will be a document outlining the argument from evidence that needs to be made in order to justify the intended uses of ACCESS 2.0 and the WIDA Screener.

Based on the agreed-upon details contained in the Framework and the TIDP, CAL will develop for its internal use detailed generative test and item specifications. These specifications, covering all details of test and item formats, including content, design, and technology specifications, will guide the development of the initial prototype assessment tasks to be researched and developed as part of this project. Research conducted on these prototypes will either confirm the specifications or inform modifications to them in an iterative fashion during the piloting stage of test development. The final version of these generative specifications will guide the development of all items needed during the field testing stage of test development.

These levels of project documentation are represented in Figure 3.

Figure 3: Levels of project documentation



## 2.5 ACCESS and ACCESS 2.0

ACCESS 2.0 and the WIDA Screener have as their foundation WIDA’s K-12 annual summative English language proficiency test, ACCESS for ELLs. ACCESS 2.0 is not, in other words, being developed from scratch, but is, rather, building on a respected and widely-used assessment. ACCESS for ELLs and ACCESS 2.0 have many features in common:

- Each item or task targets at least one of the five foundational WIDA English Language Development Standards: Social and Instructional Language (SIL), Language of Language Arts (LoLA), Language of Math (LoMa), Language of Science (LoSc), and Language of Social Studies (LoSS).
- The tests are comprised of four domain-specific subtests: Listening, Reading, Writing, and Speaking.
- Test takers are evaluated according to three performance criteria: at the discourse (linguistic complexity), sentence (language forms and conventions), and word/phrase (vocabulary usage) levels.
- A test taker’s score is mapped onto one of five language proficiency levels: “Entering,” “Emerging,” “Developing,” “Expanding,” and “Bridging.” (The “ceiling” of English language proficiency defined by the WIDA standards for assessment purposes is called “Reaching.”)
- What language ELLs will process, understand, produce, or use at each of the five defined language proficiency levels is specified in performance definitions that address desired linguistic attainments at the three levels of language analysis: discourse, sentence, and word/phrase.
- Within each combination of grade-level cluster, standard, and language domain, Model Performance Indicators (MPIs) found in or developed from the WIDA ELD Standards describe the expectations for ELLs at each of language proficiency levels 1-5 and are operationalized on the

test. The sequence of five MPIs together describes a logical progression and accumulation of skills from the lowest level of English language proficiency to the full English language proficiency required for academic success.

More information on these features of the ACCESS tests new and old can be found in *The WIDA Consortium English Language Proficiency Assessment Framework: Annual Summative and On-demand Screener*, on the WIDA ([www.wida.us](http://www.wida.us)) and ASSETS Project ([www.assetsproject.org](http://www.assetsproject.org)) websites, and in the appendix at the end of this document. Table 2 lists some ways in which ACCESS 2.0 and the WIDA Screener differ from past incarnations of the ACCESS test and indicates where the reader can find out more.

Table 2: Major features of ACCESS 2.0 (summative and screener)

Feature		Where to find more information
The summative test and screener are part of a much larger assessment system.		TIDP 2.3
Extensive documentation will be produced.		TIDP 2.4
The tests will be aligned with WIDA’s 2012 Amplification of the ELD Standards, which correspond to college and career readiness standards, including the Common Core State Standards.		Framework 5.1.3
Grade 1 will have its own test form, Grades 2-3 will comprise a new grade cluster, and Grades 4-5 will comprise a cluster.		Framework 5.2
The test will more systematically sample the spectrum of language functions that comprise academic English language proficiency.		TIDP 4.4
ACCESS 2.0 and the ACCESS 2.0 Screener will be online and computer-delivered.	The Speaking subtest will not be delivered face-to-face. Speaking prompts will be delivered by computer and student speech samples digitally recorded and, for the annual summative assessment, centrally scored.	TIDP 8
	The Listening prompts will be pre-recorded and delivered by computer.	TIDP 5.3
	The Listening and Reading subtests will be automatically scored by the computer.	TIDP 5.6, 6.6
	For scoring Writing and Speaking on the Screener, local educators will benefit from a computerized multimedia scorer training program (WIDA Screener Rater Training Program—Speaking and WIDA Screener Rater Training Program—Writing) that individualizes instruction on rating to the individual learning trajectory of the educator.	TIDP 7.7.2, 8.6.2
	A tech coordinator position will be created while the test administrator’s duties will be reduced.	Framework 8.2

### 3.0 ACCESS 2.0 at a Glance

Like ACCESS for ELLs, ACCESS 2.0 will be comprised of four subtests, one in each of the four language domains: Listening, Reading, Writing, and Speaking. There will be three separate administrations, one for Listening and Reading and one each for Writing and Speaking. For reasons explained in the section on adaptivity (4.4), the sequence of administration is: Listening and Reading in that order followed by Writing and Speaking in whatever order is logistically easier. Except for special cases—students requiring the paper-based accommodation or districts that do not yet have the necessary technology—ACCESS 2.0 will be online and computer-delivered. Students in grades 1-3, however, will take the Writing subtest entirely on paper. Students in grades 4-5 will see computer-delivered prompts for the Writing subtest but will handwrite their responses unless they are identified in advance as needing to keyboard their responses. Students in grades 6-12 will see computer-delivered prompts and keyboard their responses unless they are identified in advance as needing to handwrite their responses to writing prompts (see 7.5). All subtests of ACCESS 2.0 can be group administered and, with the exception of Writing in grades

1-3, administrations can include students in different tiers. Each subtest of ACCESS 2.0 will include an introduction and practice questions. These features of ACCESS 2.0 are summarized in Table 3.

Table 3: ACCESS 2.0 at a glance

ACCESS 2.0				
Online ACCESS 2.0			Paper-based accommodation	
TUTORIAL				
LISTENING/READING		SPEAKING	LISTENING/READING	SPEAKING
▶group administration, all grades and tiers concurrently		▶group administration, all grades and tiers concurrently	▶group administration, by tier	▶one-on-one administration
WRITING			WRITING	
(grades 1-3)	(grades 4-5)	(grades 6-12)	▶group administration, by tier	
▶group administration by tier ▶paper-based prompts and responses	▶group administration by tier ▶computer-based prompts and paper-based responses; computer-based responses allowable by students identified prior to testing	▶group administration by tier ▶computer-based prompts and responses; paper-based responses allowable by students identified prior to testing		

## 4.0 Test-level Concerns

### 4.1 Screener Functionality

As a condensed version of the ACCESS 2.0 assessment, the Screener can leverage many of the advances of this new assessment. However, the brevity of the Screener brings with it some challenges. The desire to keep the screener short must be balanced with the need for it to yield a reliable measure of a test taker's ELP.

Students at lower levels of ELP rapidly reach a ceiling on test items that they can answer, requiring brevity in the assessment before they become frustrated with it.

Students at higher ELP levels require more screener items to be administered, for multiple reasons. First, an increasing complexity in screener items is needed in order to determine the maximum proficiency of

the student. Secondly, as high proficiency may factor strongly in the decision to *not* designate a student as an ELL, it's necessary for the results to be extremely robust at the highest proficiency levels to ensure that students are not being underserved. This concern is largely absent when considering the ACCESS 2.0 summative assessment scores as one of multiple measures for determining if a student is proficient enough to exit ELL status, as the summative assessment is significantly larger in scope as compared to the Screener.

## 4.2 Design and functionality

### 4.2.1 General principles

ACCESS 2.0 will not simply be ACCESS for ELLs presented on a computer screen. Like ACCESS for ELLs, though, ACCESS 2.0 will be designed to facilitate test takers' efforts to demonstrate what they can do with the English language. Decisions regarding the look and functionality of the test, therefore, will be made with the following considerations in mind:

- **simplicity and consistency:** To minimize distraction and maximize ease of use, both the layout and the functionality of the test will be as uncomplicated and predictable as possible. The test will look uncluttered and function intuitively for the test taker.
- **student comfort and engagement:** The test will be welcoming to students and put them as much at ease as is possible within the context of a high-stakes test.
- **access to necessary content and supports:** Test takers will see or be able to easily access what they need to perform to their maximum potential on a given item or task.
- **construct fidelity:** Transferring ACCESS to the computer will not change what is being tested. Computerizing the test will involve identifying which features of the test are artifacts of the current test's paper/in-person delivery or of the computerization process, and which are construct-relevant, in order to minimize construct-irrelevant sources of differences between student performances.
- **developmental differences:** To best serve all test takers, the look and functionality of the test will differ across grade-level clusters. Careful thought will need to be devoted to the specifics of making each computer-based test age-level appropriate.
- **accessible, non-biased items:** Acknowledge the diversity of test-takers, particularly ELLs with disabilities, by developing items using the principles of Universal Design and computer-embedded supports to meet the specific access needs of students. Carefully review items for possible bias and cultural insensitivity with this diverse group of test-takers.

The general considerations above translate into more concrete principles, such as the following:

- **Use a small, fixed number of layout templates.** Navigation components will always appear in the same place on the screen. Stimulus pictures and text, item stems, and response options will appear in predictable locations, with limited variation allowed to accommodate differences in text length, number of response options, and degree of graphic support.
- **Preserve the thematic folder concept familiar from ACCESS for ELLs.** On the current ACCESS test, items (Reading and Listening) and tasks (Writing and Speaking) are presented in

thematically linked sets called “thematic folders” or just “folders.” Presenting items and tasks in this way reduces the number of unique contexts to which test takers must orient themselves.

- **Include on the screen with the question all the information students need to answer it.** Students should not have to toggle between screens as they work through an item or task. If it is necessary to look at a graph to answer a question, then the graph, the question, and the response space will appear together on one screen. If questions pertain to a reading passage, then both the passage and questions will appear on a split-screen.
- **Provide students with an indication of their progress through the test.** As students work through each domain subtest, they will want to know how much of the test remains to be completed. Test takers are accustomed to being able to gauge their progress based on the number of pages left in the test booklet.
- **Provide multiple presentation and response formats for test questions,** including a paper test alternative for the small number of students who are unable to use a computer.
- **To the greatest degree possible, display information in a flexible format** so that perceptual features such as the following can be varied as per student need:
  - The size of text, images, graphs, tables, or other visual content
  - Color used for information or emphasis
  - The volume of speech or sound
  - The speed or timing of video, animation, sound, simulations, etc. (if used).
- **Standardize the interoperability and accessibility of computer-based test items.** The test will incorporate the Accessible Portable Item Profile (APIP) interoperability framework to allow items to be tagged so they look and function the same regardless of the browser used; and so that information about a student’s access needs and accommodations are embedded in individual test items and within the test itself (e.g., zoom or speech-to-text).



Domain-specific interface and layout issues will be addressed in later sections of this document.

#### **4.2.2 Design elements**

To achieve consistency across domains and general ease of use, certain elements will appear throughout the ACCESS 2.0 test. Other elements, though they may change form from one domain to another, may occupy a standard position on the screen. Table 4 shows a *possible* set of test-wide design elements. Their numbers may be reduced or augmented as additional decisions are made about how the test should look and work.



Table 4: Test-wide design elements

ELEMENT		DESCRIPTION	
navigation bar		A test navigation bar will appear at the bottom of the screen. This bar will remain fixed throughout the test, and navigation buttons will appear or not as needed.	
navigation bar	next button		The Next button will allow students to submit their answers and proceed to the next screen.
	progress bar		The progress bar will indicate progress through a domain subtest. (Note that the status of this element is uncertain.)
	question number	Question 1	The question number will indicate the question number within a domain subtest. (Note that the status of this element is uncertain.)
item input space		Part of the screen layout will be designated space for item input (e.g. graphics, reading texts). The look and content of the item input space will be defined by domain.	
student workspace		Part of the screen layout will be designated space for student responses (e.g., selected response options, space for writing responses). The look and content of the student workspace will be defined by domain.	

### 4.2.3 Navigation

Students accustomed to being able to revisit test questions on a paper-and-pencil test may experience anxiety if they are unable to do so on a computer-delivered test. There are compelling reasons, however, for not allowing backward navigation on ACCESS 2.0. While the proposed folder-level adaptivity (see section 4.4) precludes students revisiting a folder once it has been completed, examinees could, in theory, be permitted to revisit items *within* a folder. But communicating to examinees that they may revisit items only within a folder may confuse test takers, particularly young ones or those with low English proficiency. And allowing students to go back to previous questions may also encourage them to think that they *should* go back, that information on earlier screens could help them answer the question at hand. (Each ACCESS 2.0 item, though thematically related to the others in its folder, is independent of them.) If examinees are not allowed to return to a question once it is answered, however, they will neither waste time navigating backward and forward nor agonize over whether they should.

For the reasons outlined above, a simple, one-button navigation scheme will be implemented. There will be a Next button on each screen. If a question appears on a screen, pressing the Next button will both submit the answer entered and take the test taker to the next screen. To guard against students unintentionally skipping questions, test takers will need to select a response

### Figure 4

*a.*

Theme Title

Orientation

Theme Graphic

**Next**

*b.*

[illegible]

Next

### 4.3 Accessibility and affect

### 4.3.1 Instructions and Item Components

Students taking ACCESS 2.0 should never be confused about what they are expected to do or uncertain about how to interact with the test. Every effort will be made to ensure that test

takers at all levels of English language proficiency understand the test instructions, all of which will be in English, by targeting instructions toward students at the lowest proficiency levels. To assist in this, narrated directions and practice items will be provided at the beginning of each subtest. Graphical support will be part of these directions to support students' understanding of what they are expected to do.

As shown in Table 5, when Reading is being tested, test stimuli and questions will be text only. Likewise, when Listening is being tested, test stimuli and questions will be audio only. Comprehension of written text is part of the construct of Reading, while comprehension of spoken text is part of the Listening construct. Nevertheless, graphical support may be supplied as appropriate to the targeted level of proficiency.

For Writing, the specific task prompt (i.e., the prompt to which the student's writing is a response) will be shown as text and read aloud. For Speaking, all parts of the task will be presented in both text and audio formats. This includes instructions, the task input, and the task prompt (i.e., the input to which the student's response is a rejoinder).

Table 5: Means of delivery of item/task components

Domain	Component	Means of Delivery
READING	instructions	text and audio, with graphical support (illustrations and photographs) as appropriate
	stimulus	text, with graphical support as appropriate
	question	text
	response options	text and/or graphics
LISTENING	instructions	text and audio, with graphical support as appropriate
	stimulus	audio, with graphical support as appropriate
	question	audio
	response options	graphics and/or text
WRITING	instructions	text and audio, with graphical support as appropriate
	stimulus	text and audio, with graphical support as appropriate
	task prompt	text and audio
	response	student-generated
SPEAKING	instructions	text and audio, with graphical support as appropriate
	stimulus	text and audio, with graphical support as appropriate
	model response	audio only
	task prompt	text and audio
	response	student-generated

#### 4.3.2 Virtual test administrator

Within the context of the current ACCESS for ELLs administration, adult guidance comes from three sources: (1) the test administrator who presides over the group administration of the Listening, Reading, and Writing subtests, and who is especially important as a support to students taking the Tier A test; (2) the administrator of the one-on-one Speaking subtest; and (3) the “narrator” who delivers instructions on the media-delivered Listening subtest (2013-14).

On ACCESS 2.0 the types of adult facilitators will be reduced to two. Students will still take the test in a room supervised by a test administrator, and a *virtual* test administrator will guide the test takers’ work on the computer. While the virtual test administrator (virtual TA) may differ from one subtest to another and may be heard but not seen in domains other than Speaking, it is expected that the virtual TA will be a comforting presence across the ACCESS 2.0 practice exercises and subtests.

### 4.3.3 Variation across grade levels and language proficiency levels

The format of the existing ACCESS for ELLs varies across grade-level clusters and proficiency levels. Both the look of the test and the amount of support provided changes to reflect the differing needs of test takers; both as they grow in their academic English language proficiency and as they advance in age. This same sort of purposeful variation will be carried over onto ACCESS 2.0, with the computer-delivery of the test perhaps allowing for even more tailoring. The following characteristics of the test may vary across grade-level clusters and/or language proficiency levels:

- **font/font size:** Text presented to younger students will likely appear in a larger font, and the font itself may even be different (i.e., more “child friendly”) from that used for students in the higher grades.
- **graphics:** All graphic supports—whether illustrations or photographs—will be designed to be age-appropriate, in that they will (1) show children of similar age to the test takers in situations likely familiar to test takers or (2) try to replicate to the extent possible materials students in a particular age group are likely to encounter in the classroom. This means not only that images for younger test takers will be less detailed and more “kid-friendly” than those used on the test forms for older students, but also that graphs and maps will be presented differently to students of different ages.
- **support:** Students at lower grades or language proficiency levels will likely be given more support than those who are older or more proficient. This additional support might take the form of modeling sample responses or extensive use of graphics.

## 4.4 Key academic language uses

As stated in *The WIDA Consortium English Language Proficiency Assessment Framework: Annual Summative and On-demand Screener*, one goal is for ACCESS 2.0 and the WIDA Screener to sample from the breadth of social and academic English language critical for success in today’s classrooms. Test developers will ensure that all aspects of academic English language are covered by (1) identifying key uses of academic language students should be able to either produce or comprehend, and (2) determining how to best distribute these language uses across the test forms (i.e., grade-level clusters, standards, and domains).

The notion of Key Uses of Academic Language (Key Uses) serves as a useful heuristic to describe how language is used to engage meaningfully in academic content areas at all levels of language: discourse, sentence, and word.

WIDA defines Key Uses as “overarching ‘big idea’ academic purposes, often involving more than one language function.” Key Uses typify ways in which students are expected to use language recurrently in and across academic contexts. This places the focus on using language for meaning making and what students are doing with language to accomplish various communicative activities.

The WIDA Consortium has adopted four “overarching language functions”—Explain, Argue, Recount, and Discuss—connected to the types of language reflected in new college and career readiness

standards, as well as an assortment of underlying language functions. These Key Uses will be incorporated into the generative item specifications (See Figure 4) from which ACCESS 2.0 items and tasks are produced.

## 4.5 Adaptivity

The existing ACCESS for ELLs consists of three overlapping tiers: Tier A (targeted to students at the lowest levels of English language proficiency), Tier B (targeted to students at the mid levels of English language proficiency), and Tier C (targeted to students at the highest levels of English language proficiency). As the WIDA Web site explains, “This keeps the test shorter and more appropriately targets each student’s range of language skills. What the current tier system allows ACCESS to do, folder-level adaptivity will allow ACCESS 2.0 to do even better. Note that, while the WIDA Screener will comprise fewer folders than ACCESS 2.0 itself, adaptivity will work the same on both tests.

The proposed plan for adaptivity (see sections 5.2, 6.2, 7.2, and 8.2) was conceived with three primary criteria in mind:

- **student affect**
  - All test takers, regardless of English language proficiency level, should enjoy the boost in confidence that comes from encountering test items they can do.
  - Exposure to items targeting proficiencies much higher than a student’s current level should be minimized, to avoid discouragement.
  - The test should not be longer than it needs to be, lest it fatigue students.
- **psychometric properties of the test**
  - The test must include enough items to allow for legitimate reporting of all promised subscores.
  - Scores on ACCESS 2.0 should be at least as psychometrically precise as ACCESS for ELLs.
- **coverage of the WIDA ELD standards**
  - The test must include enough items per standard to allow for the calculation of subscores by standard.
  - Roughly the same number of folders should be devoted to each standard on ACCESS 2.0 as on ACCESS for ELLs.

Information about domain-specific adaptivity mechanisms will be provided in later sections (5.2, 6.2, 7.2, 8.2) of this document, but note here that simply moving ACCESS to the computer yields some tier-related benefits. Examinees taking the computer-based test will no longer be locked into a single tier across all domains, for instance, and, at least for the majority of students (i.e., those not taking a paper-based assessment), educators will no longer need to preorder tiered tests.

### 4.5.1 Order of subtest administration

One feature of the proposed adaptivity scheme is for tier placement on the Writing and Speaking subtests to be informed by a student’s performance on the Listening and Reading subtests (which will be scored in real time). For this to be possible, of course, a student must have completed both the Listening and Reading subtests before sitting down for either the

Writing or Speaking one. The sequence of administration is: Listening and Reading in that order followed by Writing and Speaking in whatever order is logistically easier. Note that while the Listening and Reading tests need not be taken as a unit, the recommended default is to administer these subtests one after the other.

## 4.6 Video Tutorial and Interactive Practice Tutorial

While every effort will be made to make ACCESS 2.0 as user-friendly and self-explanatory as possible, students may need to be introduced to key features of the computerized test before it is administered to them. The practice exercises that will precede each subtest of ACCESS 2.0 will both expose students to the types of items or tasks they will encounter and give test takers a chance to familiarize themselves with any domain-specific functionality (recording controls for Speaking, for instance). To provide the opportunity for students to gain exposure to the test's functionality prior to taking the test operationally, however, a Video Tutorial will be developed to present sample items to students and to explain and model for students the functionality of the test environment. Additionally, an Interactive Practice Tutorial will be developed to orient students to the computer interface and give them the chance to try out their understanding of the functionality under controlled circumstances. If students know and are comfortable with how the test works, they will be able to demonstrate their current level of academic English language proficiency.

The ACCESS 2.0 Video Tutorial and Interactive Practice Tutorial

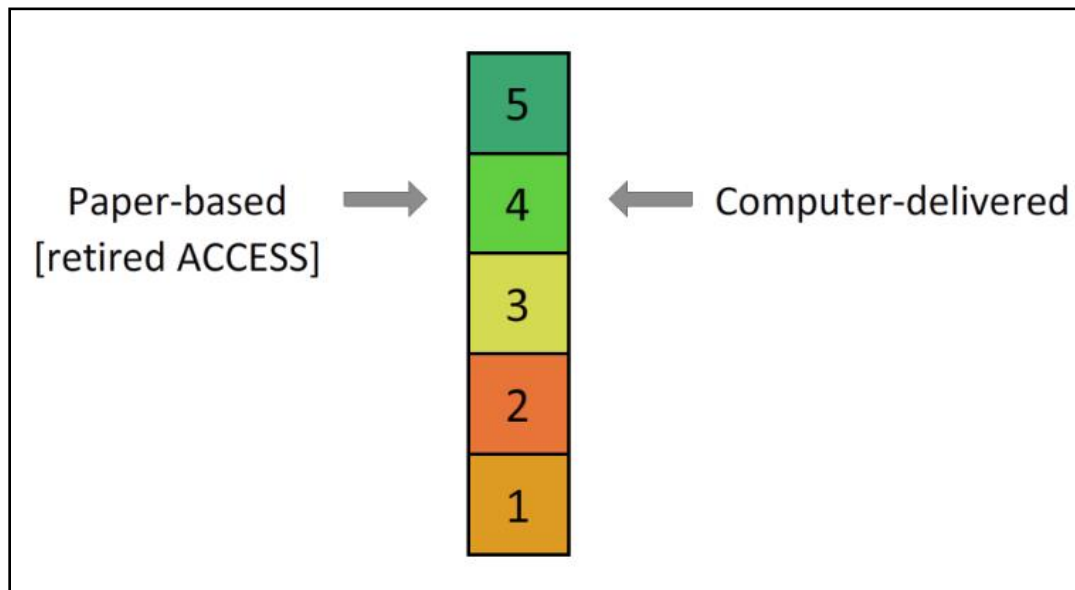
- **will introduce students to the interface, general test functionality, and the item and task types for each domain.** For example, the Video Tutorial will present items and show how a response is recorded. The Interactive Practice Tutorial will include questions that require students to record their response (used for all students in the Speaking domain).
- **may be administered to the same student more than once.** The Video Tutorial and Interactive Practice Tutorial will be available for administration throughout the school year to familiarize test-takers with the interface prior to administration of ACCESS 2.0.
- **can help schools work out the logistics of administering a computer-delivered test.** Having the Interactive Practice Tutorial as a dry run may help ensure that everything runs smoothly at the time of actual test administration.

## 4.7 Paper-based accommodation

A paper-based accommodation will be available—for both ACCESS 2.0 and the WIDA Screener—for those few individual students who may not be able to take an online test. The paper-based ACCESS 2.0 test will be primarily comprised of items from retired forms of the ACCESS for ELLs test. Since the online annual ACCESS 2.0 takes advantage of innovations offered by a technology-based assessment and undergoes an annual refreshment process, the items and tasks on the paper-based test will not always be the exact same items and tasks on the online test. Nevertheless, every effort will be made to make the two versions as parallel as possible. For example, the Speaking section of the paper-based accommodation will be based on the specifications used on ACCESS 2.0 and will include a student test booklet containing visuals from the computerized version, a playback of the audio heard in the computerized version, and, as needed, a recorder for recording student responses for later scoring.

Psychometric analyses and equating will ensure that performances on both forms will be equivalent in terms of the interpretation of the performances on the five levels described in the WIDA standards. Figure 5 illustrates this equivalence.

Figure 5: Common interpretation of the paper-based and computer-delivered versions of ACCESS 2.0



It is important to note here that there will in fact be three ways to take ACCESS 2.0:

1. **Entirely on the computer:** In this scenario, test takers see item and task prompts onscreen and input their answers directly into the computer, whether by selecting a response option for Listening and Reading, recording their spoken responses, or keyboarding their written responses.
2. **On the computer except for handwritten Writing responses:** Pre-identified students, whose keyboarding skills are not fully developed, will take the computerized test but write their responses to Writing prompts—still presented onscreen—on paper rather than keyboarding them. Students who handwrite their responses will do so in a pre-ordered, pre-labeled test booklet. (Note that all students in grades 1-3 will take a paper-based Writing subtest.)
3. **Entirely on paper:** There will be a paper-based test form for all domains available for students who need it as an accommodation.

*[N.B. More information about other accommodations will be added to this document in reflection of the work of the ASSETS project Accommodations, Accessibility, and Equity subcommittee.]*



## 4.8 Reporting

### 4.8.1 Scores reported

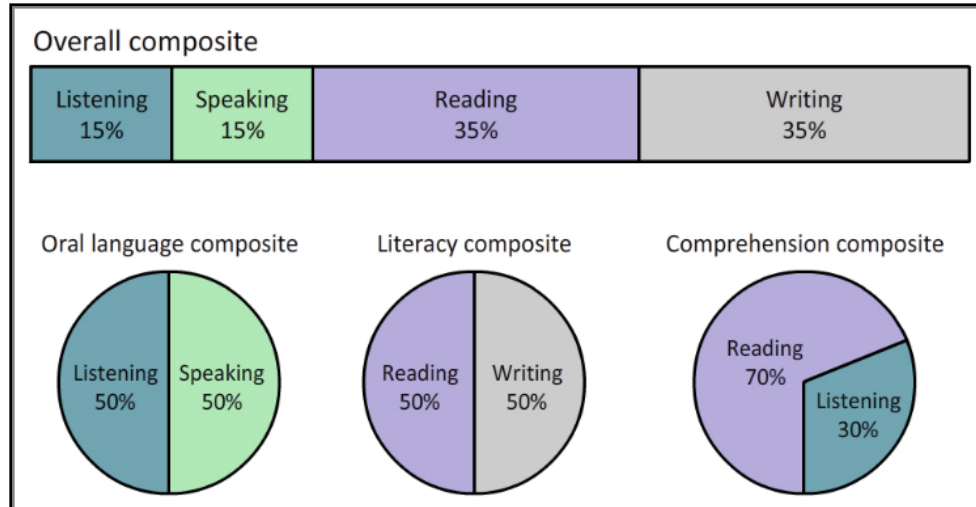
#### 4.8.1.1 ACCESS 2.0

One goal of the ASSETS Project is to improve the ACCESS 2.0 score reports in terms of the usefulness of information provided. Psychometric work, in combination with careful descriptions of the domains and standards, will ensure that all scores are valid and reliable. Through the work of the Reporting subcommittee of the ASSETS Project, stakeholders will be consulted to determine what additional information would be helpful to add to the score reports, or how scores could be more clearly communicated.

Student data from ACCESS 2.0 will therefore include but not necessarily be limited to:

- **scale scores by domain** (Listening, Reading, Writing, and Speaking) on a K-12 vertically aligned scale;
- grade-level specific **interpretive proficiencies by domain**, which characterize a student's performance in terms of the proficiency levels 1 through 5 defined in the WIDA standards; and
- **composite scores** as diagrammed in Figure 6.

Figure 6: Composite scores



*[N.B. While the timeframe for the delivery of official score reports including all of the abovementioned scores will depend on the pace of the centralized scoring of Writing and Speaking tasks, the Reporting subcommittee will explore the possibility of releasing unofficial results from the computer-scored Reading and Listening subtests earlier for use in local decision-making.]*

#### 4.8.1.2 The WIDA Screener

The WIDA Screener will be shorter than ACCESS 2.0 (for details, see sections 5.7, 6.7, 7.7, and 8.6), and the overall English language proficiency level score will be the most psychometrically valid piece of data derivable from a student's performance on the test. Initial proficiency-level scores for each domain may prove helpful in determining the extent of students' English language support needs in each domain.

### 4.8.2 Score reports

#### 4.8.2.1 ACCESS 2.0

Score reports generated for ACCESS 2.0 will build on WIDA's seven-year experience with delivering meaningful, uniform score reports customized to the needs of various stakeholders. Although the specific score reports to be produced will be determined by the WIDA Consortium member states, it is expected that score reports will target audiences similar to those receiving ACCESS for ELLs score reports. The format of the reports may be altered to more clearly communicate to individual groups of stakeholders the information they most need and want.

Currently, WIDA provides for the ACCESS assessment the reports listed in Table 6.

Table 6: Score reports

Report	Description
<b>parent/guardian</b>	<ul style="list-style-type: none"><li>• test results presented visually and numerically</li><li>• includes domain, oral language, literacy, comprehension, composite English language proficiency level and scale scores</li></ul>
<b>teacher</b>	<ul style="list-style-type: none"><li>• more detailed than parent/guardian report</li><li>• includes scale scores for all domains and combinations of domains and raw scores for each of the WIDA ELD standards</li></ul>
<b>student roster</b>	<ul style="list-style-type: none"><li>• for teachers and administrators</li><li>• gives overview of English language proficiency levels and scale scores for all domains and composite scores for ELLs in a school</li></ul>
<b>school frequency</b>	<ul style="list-style-type: none"><li>• for teachers and administrators</li><li>• shows distribution of ELLs according to their language proficiency levels for each domain and combination of domains in a school</li></ul>
<b>district frequency</b>	<ul style="list-style-type: none"><li>• provides the same information as the school frequency report across an entire district</li></ul>

In collaboration with SEAs and LEAs, WIDA is already providing translations of the parent/guardian report in more than 30 languages and will continue this practice with the ACCESS 2.0 assessments.

A Reporting subcommittee will also make recommendations on whether the 100-600 scale used for ACCESS for ELLs will be retained or whether a new scale will be instituted to differentiate ACCESS scores from ACCESS 2.0 scores. They will also consider different ways of presenting such information as the confidence bands associated with scale scores and the relationship between scale scores and proficiencies. Input from focus groups, psychometricians, and design professionals will help ensure that score reports are maximally comprehensible and useful for their intended audiences.

#### **4.8.2.2 The WIDA Screener**

For the WIDA Screener, a technology-based administrator interface will be developed to streamline and standardize the process of reporting results. The local program will enter a student's Speaking and Writing scores into the computer, where they will be combined with the Listening and Reading scores already stored there. The computer will do the number-crunching and then generate a printable score report. Besides easing the burden on the test administrator, digitizing results in this way will allow for easy aggregation of student data and monitoring of screener use.

## **5.0 Subtest Specifics: Listening**

### **5.1 Composition of the Listening subtest**

Figure 8 shows the proposed structure of the Listening subtest. There are three major components:

- Panels: Numbered columns headed by a standard abbreviation (SIL, LoLA, e.g. ; see Section 2.4 for their definitions)
- Folders: Stacks of rectangles represent a thematic folder
- Items: Small rectangles containing a number which indicate the level of the MPI found in or derived from the WIDA ELD standards targeted by the item

Though all but the first two panels include more than one folder, a test taker will be administered only one folder per panel, as appropriate for their ELP level. (Details about how this is determined can be found in the next section on Adaptivity.) Readers familiar with ACCESS will note that the structure of the item pool mirrors that of the three tiered forms within a grade-level cluster on the current test. This relationship is highlighted by marking rows as "A", "B", and "C."

All students will take one folder targeting Social and Instructional Language in each of the first two panels. (entry folders) After the first two panels, students will take one folder targeting each of the academic standards LoLA, LoMA, LoSS, and LoSC. (Panels 3-6) Students at higher language proficiency levels will see an additional folder in each of the standards of LoLA and LoMa. (Panels 7 & 8) Depending on a test taker's language proficiency, the Listening subtest of ACCESS 2.0 will consist of between 18 and 24 selected response items, grouped in either six or eight thematic folders.

Figure 7: The structure of the Listening subtest



## 5.2 Adaptivity

All test takers will begin the Listening subtest of ACCESS 2.0 with two entry folders, both targeting Social and Instructional Language (SIL). The test taker's performance on the six items in the entry folders will determine which of the three leveled LoLA folders in Panel 3 s/he is administered. Throughout the test, an examinee's underlying measure of ability will be recalculated after each folder s/he completes, with the tier of the next folder to be administered chosen accordingly. Students will not all see the same folders, but the order of the panels will be invariant across test takers. All students, in other words, will have the standards targeted in the left-to-right order shown in Figure 7. Figure 8 shows one path a student might take through the Listening subtest. Note that the test taker always begins a folder with the least challenging question and works up to (and exits the folder after) the most challenging.

[illegible]

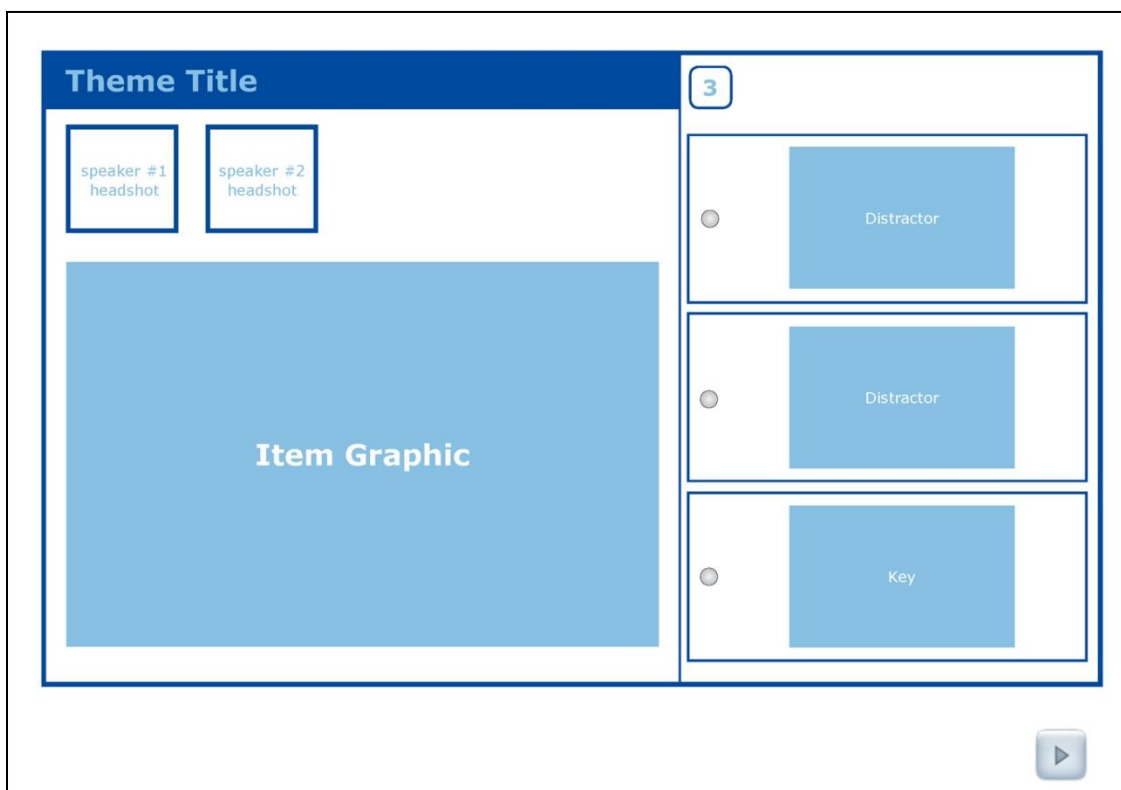
### 5.3 Item presentation

The auditory stimuli on the Listening test will also be supported visually. The introductory screen of each folder will include a graphic that establishes the context of the speech students will hear in the items that follow. The graphic will show who will be speaking and the situation in which the speech will occur.

## 5.4 Domain-specific design elements

As indicated in the preceding section, on the Listening subtest of ACCESS 2.0, students will hear not only monologues but also dialogues. To help students keep track of who is talking as they listen to multiple speakers interact, each question screen will display headshots of the speakers involved in that item. Test takers will have seen the speakers in the introductory graphic used to set the context for the whole folder, and thus should be able to recognize the speakers from headshots on subsequent screens. Figure 9 shows how representation of the speakers will be integrated into the layout of a Listening screen.

Figure 9: A possible look for the Listening subtest



## 5.5 Student response

Students will respond to all items on the Listening subtest by selecting an option from among three choices.

*[N.B. As stated in the grant proposal, alternatives to multiple choice—more technology-enhanced and performance-based item types—will be researched and developed for the lower stakes, optional interim assessments. They will be introduced onto the Reading and Listening subtests of ACCESS 2.0 and the ACCESS 2.0 Screener as appropriate, but not until the post-grant period. This delay will allow their usefulness and stability to have been researched and demonstrated on the interim assessments.]*

## 5.6 Scoring

Student responses on the Listening subtest will be automatically and immediately scored by the computer.

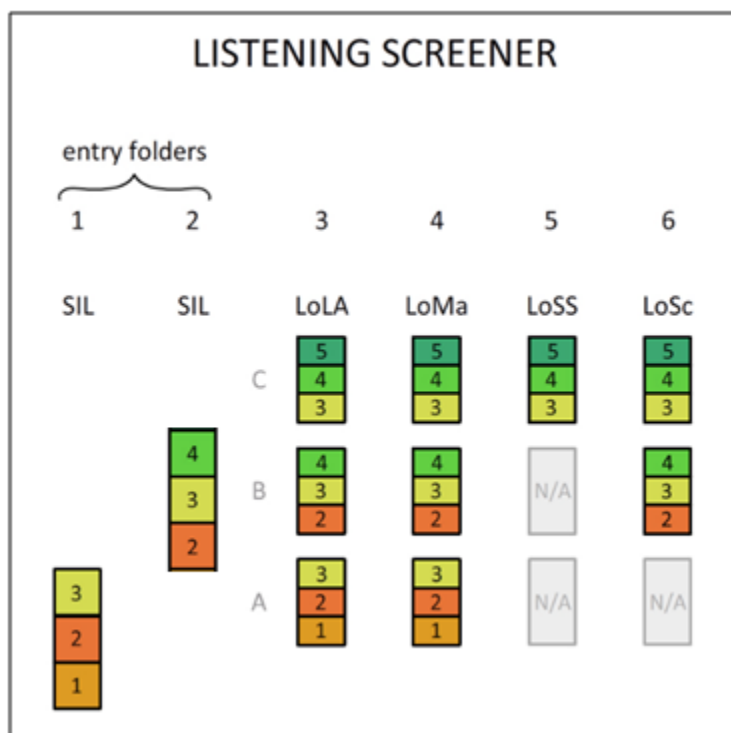
*[N.B. Score reports may not be available for the summative assessment, however, until the centralized scoring of the Writing and Speaking tasks has been completed.]*

## 5.7 Listening screener

The Listening screener is envisioned as a shortened version of the ACCESS 2.0 Listening subtest. Adaptivity will work on the screener as it does on the summative subtest. Students who demonstrate high levels of English language proficiency on the Listening screener will take a total of six folders, including one for each standard. Students who track at the mid or low proficiency levels, however, will finish the screener after seeing a reduced number of folders.

Note that while students taking primarily Tier B folders will not be exposed to all standards on the Listening screener, the standard they miss—Language of Social Studies—will be covered on the Reading screener (see Section 6.7). Figure 10 shows the structure of the Listening screener.

Figure 10: The structure of the Listening screener







### 6.3 Item presentation

On the Reading subtest, texts will appear on the computer screen. Care will be taken to alleviate the strain of reading on the computer screen as much as possible. For each thematic folder, the folder introduction screen will provide visual support for the context or the type of text-based material (e.g., book, newspaper, Web site) from which the text is drawn. Cognitive labs and pilot testing have informed decisions about how to design these folder introduction screens, lay out Reading passages, and equip students of different age groups to navigate them.

## 6.4 Domain-specific design elements

The most distinguishing feature of the Reading subtest is the relatively large amount of text that must be presented to test takers. Text (and visuals as appropriate) will be presented on the left-hand side of the screen, occupying 60% of the screen width. Where possible, text and visual elements will be arranged to avoid the need to scroll the text, particularly for the youngest learners.

The question and answer options will be presented in a single column on the right-hand side of the screen, occupying about 40% of the screen width. The need to scroll through answer choices will be avoided. The basic layout of text, question and answer options is shown in Figure 13.

**Figure 3: A possible look for the Reading subtest**

[illegible]

## 6.5 Student response

Students will respond to all items on the Reading subtest by selecting an option from among three or four choices, as is done on the current ACCESS for ELLs.

*[N.B. As stated in the grant proposal, alternatives to multiple choice—more technology-enhanced and performance-based item types—will be researched and developed for the lower stakes, optional interim assessments. They will be introduced onto the Reading and Listening subtests of ACCESS 2.0 and the WIDA Screener as appropriate, but not until the post-grant period. This delay will allow their usefulness and stability to have been researched and demonstrated on the interim assessments.]*

## 6.6 Scoring

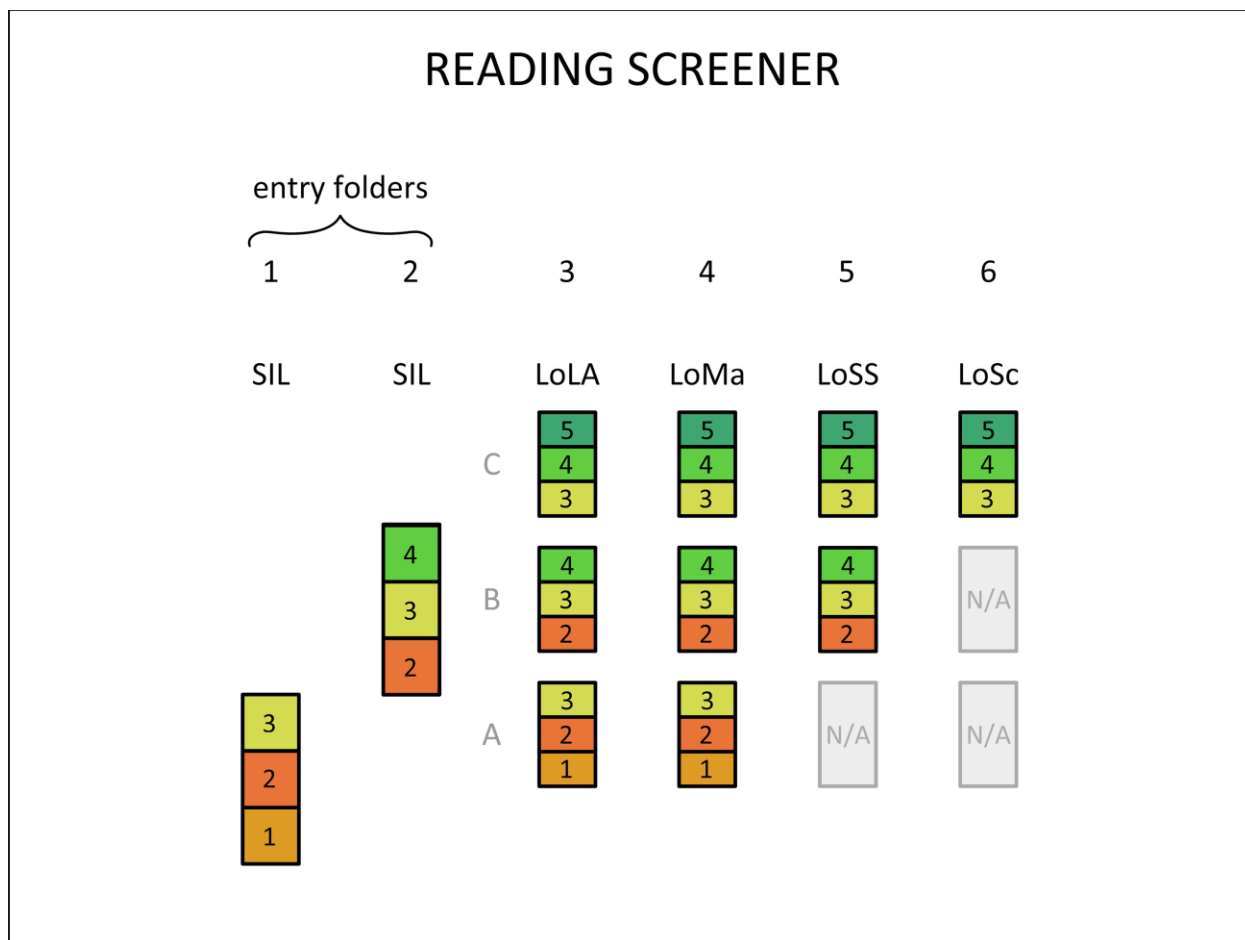
Student responses on the Reading subtest will be automatically and immediately scored by the computer.

*[N.B. Score reports may not be available for the summative assessment, however, until the centralized scoring of the Writing and Speaking tasks has been completed.]*

## 6.7 Reading screener

The Reading screener is structured similarly to the Listening screener (see section 5.7). Note that students tracking at the mid levels of language proficiency will see a Language of Social Studies folder on the Reading screener, to compensate for the lack of such a folder on the Listening screener.

Figure 4: The structure of the Reading screener



## 7.0 Subtest Specifics: Writing

### 7.1 Composition of the Writing subtest

The Writing subtest will comprise three tasks. On the Tier B/C form, the third task will be an extended task, as shown in Figure 15. The way in which the standards are targeted by these tasks will vary across grade levels and will be spelled out in the generative item specifications. (See Figure 4)

### 7.2 Tier placement

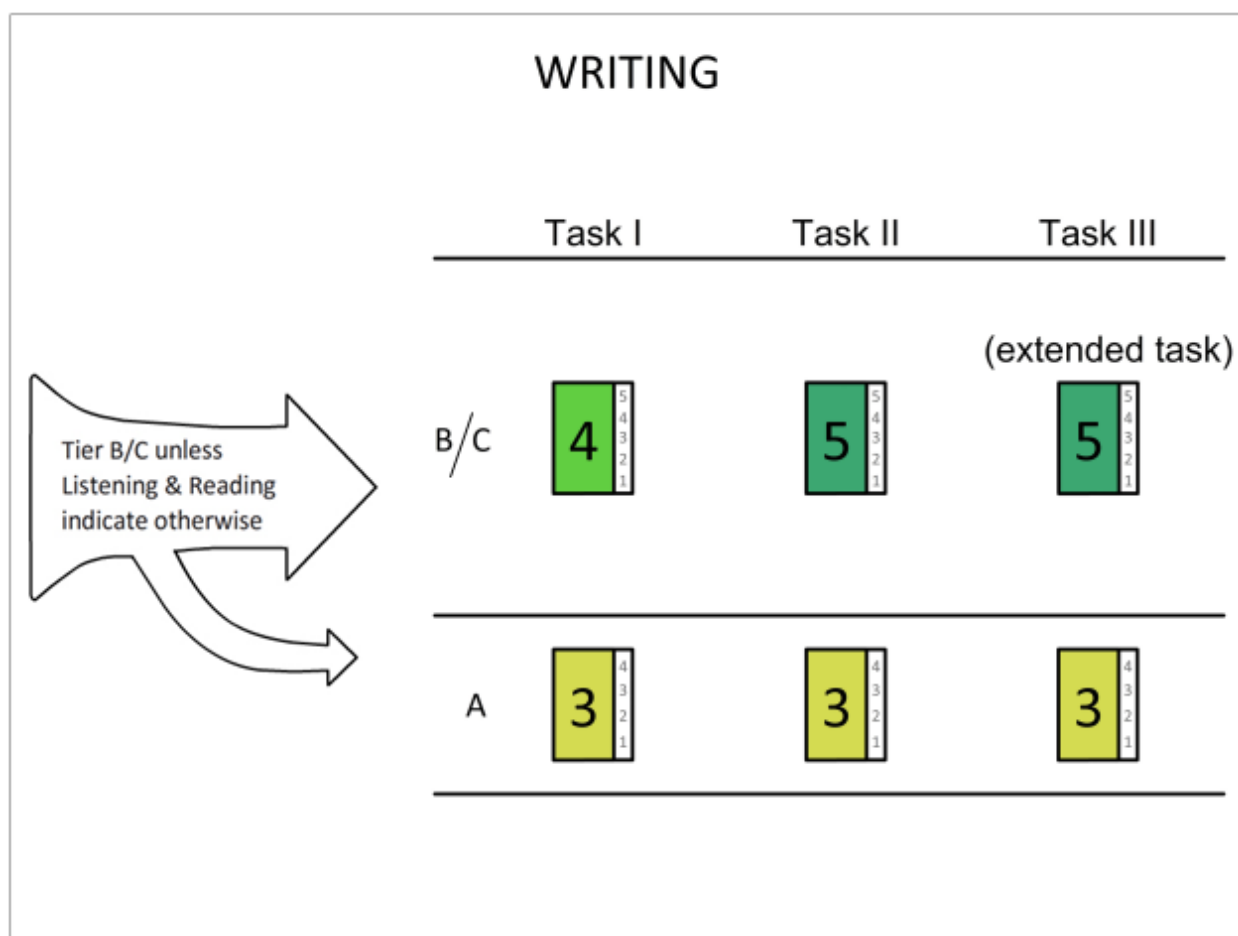
Tier placement into the Tier A or Tier B/C Writing subtest will work differently depending on whether students are taking a paper-based or computer-delivered test.

All examinees in grades 1-3 will take the Writing subtest on paper, grouped by tier (A or B/C), as will any older student judged to be better served by the paper-based accommodation for the entire test. Tiered paper-based tests will have to be ordered in advance, with tiering decisions made locally, similarly to how they are made for the current ACCESS test.

Tier placement for the computer-based Writing subtest, however, will be informed by student performance on the Listening and Reading subtests. While most test takers will be routed by default into the Tier B/C Writing subtest, a smaller subset of test takers is expected to be placed in Tier A on the basis of their scores in the receptive domains of Listening and Reading. Data from the current operational ACCESS program will help inform when it is appropriate to route a student to Tier A rather than Tier B/C. Once a test taker is placed in a tier, s/he remains in that tier for the duration of the subtest.

In Figure 15, the bold numbers within the colored rectangles indicate the MPIs the tasks are written to target. The small gray numbers along the right of each rectangle indicate the range of scores attainable on that task. These serve as a reminder that although a test taker is routed into and stays within a tier, there are a wide range of scores within it, allowing students to demonstrate their maximum proficiency.

Figure 5: The structure of the Writing subtest



### 7.3 Task presentation

Except for examinees in grades 1-3 and those taking the paper-based accommodation, writing prompts will appear on the computer screen. While it may not be practicable to incorporate animations or video

into the initial version of ACCESS 2.0, the computer’s capability to bring graphics to life may be used on future iterations of the test to increase the chances of eliciting rich writing samples from students. In the spirit of providing test takers with maximal support and making every provision to ensure that they have the opportunity to demonstrate the full extent of their English language proficiency, some modeling may be used to help make task expectations as clear as possible, particularly for test takers currently at the lowest levels of English language proficiency. The first of a series of questions might be filled in already, for example, or a sentence starter might be provided.

## 7.4 Domain-specific design elements

For ease of reading, the stimulus and student response boxes on the Writing subtest of ACCESS 2.0 will each occupy approximately half of the width of the screen. This adheres to the ACCESS 2.0 principle of keeping each item or task restricted to a single screen. Figure 16 shows a possible look for the Writing subtest.

Figure 6: A possible look for the Writing subtest

The interface is designed for a writing task. It features a dark blue header on the left with the text "Theme Title". Below this header is a large light blue rectangular area labeled "Item Graphic". To the right of this area is a white rectangular area. At the top of this white area is a small blue box with the number "3" in white, followed by the text "Task statement/question" in blue. Below this is a large white rectangular area with the text "Type here" in grey. At the bottom right of the entire interface is a small blue square button with a white right-pointing triangle.

## 7.5 Student response

Examinees in grades 1-3 and those taking the paper-based accommodation will plan and write their responses in paper test booklets. Before they actually begin to write, test takers in Tier B/C will be given

scrap paper to use for pre-writing activities, if they choose to use it. Writing on this scrap paper is not scored, and the paper will be destroyed after test administration.

As noted in Section 3, some students in grades 4-5 and most students in grades 6-12 taking the computerized version of ACCESS 2.0 will keyboard their responses, while most students in grades 4-5 and some students in grades 6-12 will write them by hand. Students responding on paper will have been pre-identified and will write their responses in pre-ordered and pre-labeled writing student response booklets. A mechanism will be built into the Writing test to ensure that it is clear to the centralized scoring center where—on the computer or in a test booklet—to find each test taker’s response.

To determine what editing tools should be made available to students who compose their Writing responses on the computer, WIDA staff will monitor the work of the content consortia Partnership for the Assessment of Readiness for College and Careers (PARCC) and Smarter Balanced Assessment Consortium (SBAC) and follow their lead. If students will have such word processing functionality as cut-and-paste and block deleting at their disposal on content tests, ACCESS 2.0 should provide these capabilities as well.

Table 7 summarizes the delivery modes used for various components of the Writing subtest.

**Table 7: Delivery modes on the Writing subtest**

Grade range	Prompt presentation	Scripting	Planning space	Student response
<b>1-3</b>	paper test booklet	live TA	in test booklet	paper test booklet only
<b>4-5</b>	onscreen	virtual and live TA	scrap paper	writing response booklet (or computer for pre-identified students)
<b>6-12</b>	onscreen	virtual and live TA	scrap paper	computer (or writing response booklet for pre-identified students)

## 7.6 Scoring

Whether they are keyboarded directly into the computer or written on paper and then scanned, responses to ACCESS 2.0 Writing tasks will be centrally scored by trained raters. For the field testing during the grant period, student writing will be sent to MetriTech, Inc., where highly-trained raters are already familiar with the WIDA rubric for scoring Writing. The Writing rubric currently used for ACCESS for ELLs will be updated to reflect the 2012 amplification of the WIDA ELD Standards and performance definitions.

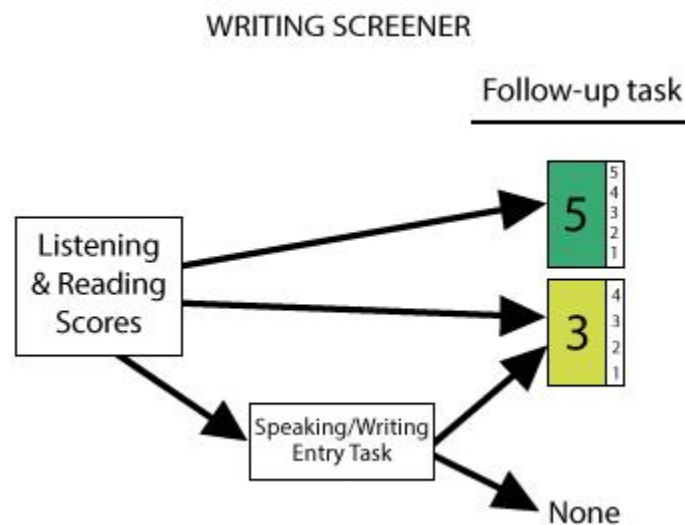
## 7.7 Writing screener

### 7.7.1 Composition of the Writing screener

In order to obtain a reliable read on a student’s writing ability, it is necessary to elicit a sizable writing sample. The Writing screener, therefore, will include an extended task for test takers

with higher English language proficiency. Test takers will be routed to either the extended task, a shorter task, or a Speaking/Writing entry task based on their performance on the Reading and Listening sections of the WIDA Screener as shown in Figure 17. If a test taker's score on the Listening and Reading sections indicates that he or she appears to not yet have developed any English language proficiency, he or she will be administered a quick Speaking/Writing entry task. The short, integrated Speaking/Writing entry task, written to its own specifications, will serve as a check on whether the test taker has any proficiency in English in the productive skills. If it becomes apparent during the entry task that a student cannot write in English at all, he or she will exit the Writing screener subsection without being administered a follow-up writing task. Test takers who have been routed to the entry task and display some ability to write in English will be routed to the shorter Writing follow-up task.

Figure 7: The structure of the Writing screener



### 7.7.2 Scoring the Writing screener

The Writing portion of the WIDA Screener will be locally scored. Materials will be developed as part of the ASSETS grant to guide the local training of raters, whether teachers, administrators, or external hires. A multimedia scorer training program will provide trainee-raters with ample opportunities to practice and to calibrate themselves to a common scale. Specifically, the training program will incorporate the following: (a) interactive, technologically supported training; (b) content-based quizzes that provide diagnostic feedback as a rater learns material; (c) scoring quizzes that provide full justifications for scores and help explain to trainees how to re-calibrate to the scale; (d) computer-adaptive scoring practice that focuses rater-trainees' attention on aspects of scoring they find difficult, using an underlying pool of pre-rated samples, helping raters internalize the criteria; and (e) a digitized library of pre-rated responses that can be used as benchmarks against which to compare writing performances.

Local scoring of written responses on the ACCESS 2.0 Screener will be facilitated through a computerized interface called the *ACCESS 2.0 Screener Rater Training Program—Writing*. This interface provides the rater with ready access to a test taker’s written responses (if digitized). Raters can also view the task instructions, sample scored responses (benchmarks), and tips on scoring, and can easily input ratings and comments.

## **8.0 Subtest Specifics: Speaking**

### **8.1 Composition of the Speaking subtest**

For each grade cluster, there will be three forms of the test: Pre-A, Tier A, and Tier B/C. The test forms address overlapping proficiency levels, 1, 1-3, and 3-5, respectively. Tier A and Tier B/C forms of the Speaking subtest of ACCESS 2.0 will comprise three folders, each based on one or two standards. Each of these folders will consist of two tasks related to a common theme. Students taking the Tier A or Tier B/C form of the test will complete a total of six tasks. Beginning proficiency students will be administered a Pre-A form of the test which contains three folders, each with one task.

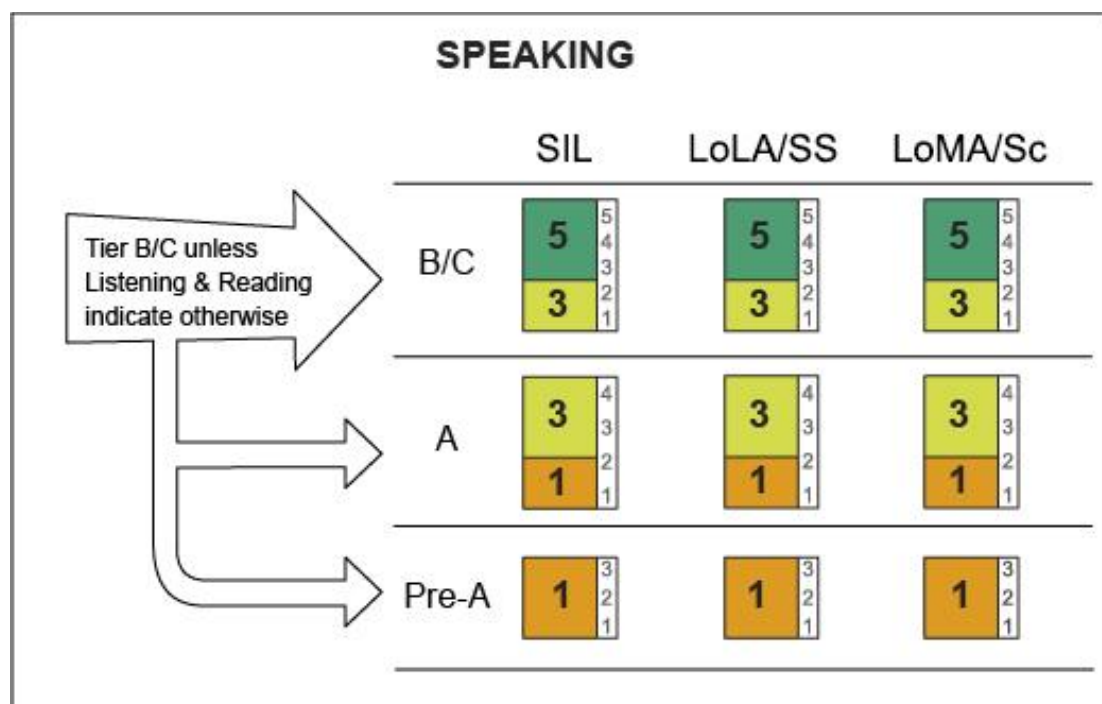
### **8.2 Tier placement**

As with Writing, tier placement for the Speaking subtest will be informed by student performance on the Listening and Reading subtests, as indicated in Figure 18. Most test takers will be administered the Tier B/C test, though some students may be routed into Tier A if their performance on the Listening and Reading subtests indicates that this is warranted. Students whose scores on the Listening and Reading subtests indicate that they are at the very beginning stages of English language development will be routed into a Pre-A form of the test, which contains three Level 1 tasks.

Although a student stays within the assigned tier for the duration of the subtest, it is important to note that confining a test taker to a particular tier is not restrictive. The tiers overlap, and tasks are constructed such that students can demonstrate higher than the targeted level of language proficiency. As in the Writing diagram, the small gray numbers along the right of the rectangles in Figure 18 indicate the scores attainable on a given task.



Figure 8: The structure of the Speaking subtest



### 8.3 Task presentation and student response

As shown in Figure 18, each test form will consist of three folders, one for SIL, one for LoLA and LoSS, and one for LoMa and LoSc. Tier A and B/C folders on the Speaking subtest will contain two tasks. Pre-A test folders will contain one task. Each task will follow the same basic structure and will be designed to support and scaffold student responses using the computer interface. Because it is important for students to speak to an audience, a virtual test administrator (virtual TA) will guide students through the Speaking tasks. A photograph of the virtual TA will be visible on the screen. The virtual TA will serve as both a narrator and as a virtual interlocutor for the student. A photograph of a model student will also appear on the screen throughout the Speaking subtest. The model student will respond to questions from the virtual TA and demonstrate the language expectations of tasks. Both the virtual TA and the model student will be shown wearing computer headsets.

As students enter a Speaking folder, they will first see a folder introduction screen. The purpose of this screen is to orient the student to the topic and contents of the folder. Tasks within a folder will be based on the same theme. The folder introduction screen will include a theme graphic and a short description of the general theme, which the student will both hear narrated by the virtual TA and see in text on the screen.

After the folder introduction, students will begin a task. Each task includes three main parts: task input, a model response, and the student response. Table 8 describes the parts of a Speaking task.

Table 8: Parts of a Speaking task

Part	Description
<b>task input</b>	audio, text and graphic support that introduces a student to the topic of the task and provides input about the task theme
<b>model response</b>	model student response to a model prompt that is the same or similar to the task prompt
<b>student response</b>	student response to the task prompt

Within a task, students will first see task input. The task input will include a task graphic (if applicable) and information about the task topic, which the student will hear narrated by the virtual TA and see in text on the screen. If present, the task graphic will remain on the screen while the student responds.

After the introduction screen, the student will hear a model student response. The purpose of the model response is to demonstrate to students the type of response (e.g., quantity of language and level of detail) that is expected. As stated earlier, a photograph of a student wearing a computer headset will appear on the screen throughout the Speaking subtest to represent the model student. The student will hear the virtual TA administer the model prompt (which will be the same as or similar to the task prompt to which the student will later respond). Next, the student will hear a model response. The model response will only be presented by audio and will not appear in text. In cases where the model prompt is identical to the task prompt to which the test taker will respond, the model response will be presented so as not to inhibit test takers' ability to provide an original response. At times, the test taker may hear only the beginning of the model response.

In the third part of a Speaking task, the student will respond to a task prompt. In some cases, the student may hear additional task input after the model student responds to a sample task prompt. The student will use an audio recording mechanism (described in section 8.4) to record responses to a Speaking task. The audio recording mechanism is a part of the test delivery interface. Instructions reminding the student how to record answers will appear on the screen. After hearing a task prompt, students may record their answers at any time.

Students will then respond to the task prompt. A speech indicator on the audio panel will indicate to students that the recording is working. The expected length of a student response will vary by task level, with higher level tasks requiring more extended discourse (and thus more time to produce). Students may press the stop button when they are done speaking; otherwise the program will automatically stop recording after a certain amount of time has elapsed. Students will then move on to the next task or folder.

## 8.4 Domain-specific design elements

During the Speaking subtest, students will see, listen to, and respond to tasks via the computer test interface. It is important to note that the Speaking test requires examinees to have a headset with a microphone. From a programming perspective, the Speaking test also necessitates the use of audio capture and audio file storage functionalities.

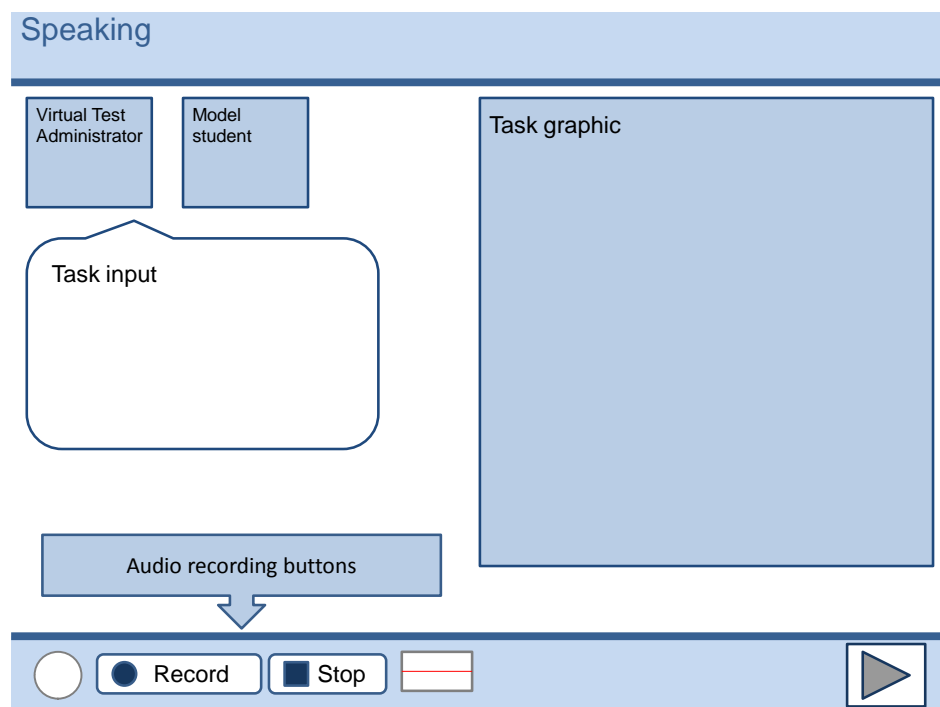
The Speaking subtest will require several domain-specific elements to address the technological demands of this section (i.e., microphone and audio recording) and to facilitate speaking into a computer. A mechanism for students to record their responses is the primary design element that will be unique to the Speaking subtest. This audio recording mechanism is part of the computer delivery interface. While the design of features may vary across computer delivery platforms, certain features are necessary in order for the Speaking subtest to function properly. A task timer will help students know when their turn to speak is almost over by visually representing time elapsing (e.g., through an indicator disappearing as time runs out). Table 9 summarizes required features of the audio control panel.

**Table 9: Speaking-specific design elements**

DESCRIPTION		
The audio recording mechanism is a fixed part of the computer delivery interface and will appear in the navigation panel throughout the Speaking subtest.		
recording mechanism	record/stop button	The student will click a button to begin recording. After the student is done speaking, the student will click a button to stop the recording.
	speech indicator	An onscreen indicator will let students know that the microphone is registering their speech and that it is being recorded.
	task timer	A visual indicator that will let students know when their time to speak is almost over

Figure 19 shows a possible look for the Speaking subtest, incorporating both the audio control panel and a representation of the virtual test administrator.

**Figure 9: A possible look for the Speaking subtest**



## 8.5 Scoring

Speaking test responses will be centrally scored. During administration, responses will be captured, saved electronically, and uploaded to the secure scoring system. For the field testing during the grant period, the digitized student oral responses will be sent to MetriTech, Inc.

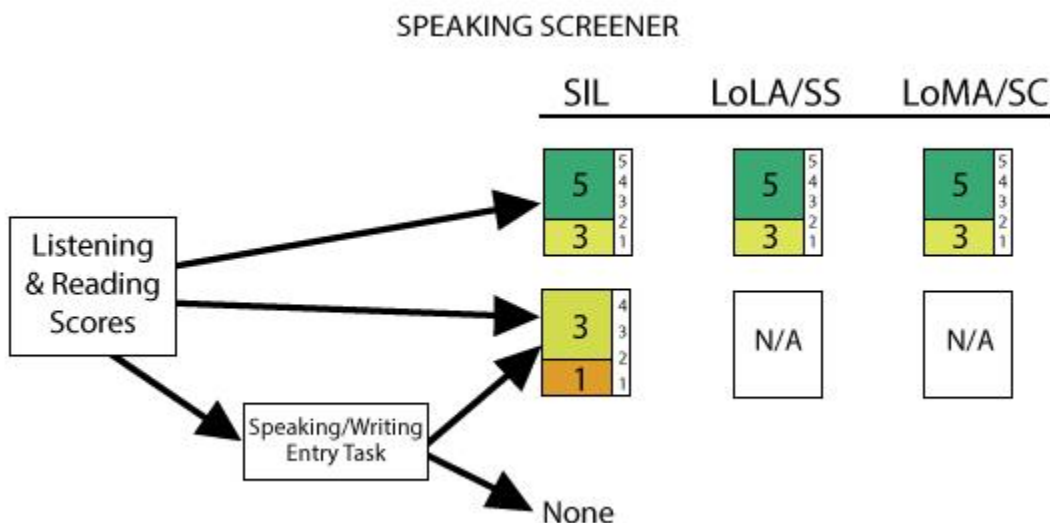
ACCESS 2.0 test development includes the creation of a new rubric for the Speaking test. This rubric will be modeled on the existing ACCESS Writing rubric. Rather than being scored dichotomously as to whether a response meets or approaches task level expectations (as in the current ACCESS), each student response will be scored across the range of proficiency levels presented on the rubric. Scores will be assigned at the folder rather than the task level. Raters may assign plus or minus scores to indicate the relative strength of the student's response at that level. In this way, the new rubric and scoring procedure will allow for the collection of more nuanced information about oral performance.

## 8.6 Speaking screener

### 8.6.1 Composition of the Speaking screener

As with the Writing screener (see 7.7.1), there are two tiers for the Speaking screener, corresponding to Tier A and Tier B/C on the summative assessment. Based on his or her score on the Listening and Reading sections, the test taker will be routed either to Tier B/C, Tier A, or to the quick integrated Speaking/Writing entry task (see 7.7.1). If it becomes apparent during the entry task that the test taker has not yet developed any speaking proficiency in English, he or she will not be administered any further speaking tasks. Test takers who have been routed to the Speaking/Writing screener and display some speaking proficiency in English will be routed to the Tier A speaking task. Figure 20 shows the structure of the Speaking screener.

Figure 10: The structure of the Speaking screener



### 8.6.2 Scoring the Speaking screener

The Speaking portion of the WIDA Screener will be locally scored by educators trained using a computer-based scorer training program called the *WIDA Screener Rater Training Program—Speaking*, based on the Multimedia Rater Training Program (MRTP) developed by CAL. The scorer training will incorporate the following: (a) interactive, technologically supported training, (b) content-based quizzes that provide diagnostic feedback as a rater learns material, (c) scoring quizzes that provide full justifications for scores, (d) scoring practice (using an underlying pool of pre-rated samples) that focuses rater-trainees’ attention on aspects of scoring they find difficult and helps rater-trainees internalize the scoring criteria, and (e) a digitized library of pre-rated responses that can be used as benchmarks against which to compare student performances.

In addition, local rating of oral responses on the WIDA Screener will be facilitated through a computerized interface adapted from CAL’s Computerized Oral Proficiency Instrument. This interface will provide the rater with ready access to a test-taker’s oral responses, with full playback/pause/repeat functionality. Raters can also view the task instructions, sample rated responses (benchmarks), and tips on rating, and can easily input ratings and comments.

## 9.0 Next Steps

This document is not an end in itself, but a means toward making ACCESS 2.0 and the ACCESS WIDA operational by 2015-16. Table 10 captures the phases of development and has been informed by the pilot test and first round of field testing. It also includes and the second round of field testing and the operational roll-out.

Table 10: Phases of development

Year	Phase of Development	Project Activities
------	----------------------	--------------------

2011-12	Initial development	<ul style="list-style-type: none"> <li>• create initial overall test design and development plan and initial test and item specifications, with input and consensus from stakeholders</li> <li>• develop and review item pool for piloting and assemble pilot-test forms</li> <li>• develop ancillary materials for pilot test (including administrator/scorer materials)</li> </ul>
2012-13	Pilot testing	<ul style="list-style-type: none"> <li>• pilot items through iterative cognitive labs and tryouts</li> <li>• conduct analyses on data collected through cognitive labs and tryouts</li> <li>• revise overall test design and development plan, including generative test and item specifications (as needed)</li> </ul>
2013-14	Field testing	<ul style="list-style-type: none"> <li>• develop and review item pool for field test</li> <li>• assemble field-test forms</li> <li>• develop ancillary materials for field test (including administrator/scorer materials)</li> <li>• conduct field test</li> <li>• score field test</li> <li>• conduct field-test item analyses and reliability studies</li> </ul>
2014-15	Additional field testing & Pre-operationalization	<ul style="list-style-type: none"> <li>• conduct, score, and analyze results of field tests</li> <li>• conduct standard setting</li> <li>• assemble final operational test forms</li> <li>• develop final score reports and score reporting system</li> <li>• develop ancillary and training materials for operational test</li> <li>• conduct additional reliability and validation studies</li> <li>• finalize plan for continual monitoring and evaluation of operational testing program</li> </ul>

## Appendix: Understanding the current ACCESS for ELLs Assessment

### A1.1 Introduction

As mentioned in sections 1.2 and 2.4, this appendix has been included with the *Test and Item Design Plan* to give readers unfamiliar with ACCESS for ELLs background about ACCESS 2.0's illustrious precursor.

ACCESS for ELLs (Assessing Comprehension and Communication in English State-to-State for English Language Learners) is a secure, large-scale English language proficiency assessment given to students in grades K-12<sup>2</sup> who have been identified as English language learners (ELLs). It is given annually in WIDA Consortium member states to monitor students' progress in acquiring academic English. This document explains the structure of the test, including the grade-level clusters and tiers.

### A1.2 Test structure

The ACCESS for ELLs test battery is a collection of assessment instruments administered to all ELL students across grades K-12. It first and foremost operationalizes the English language development standards<sup>3</sup> that form the core of the WIDA Consortium's approach to serving the needs of English language learners. These standards incorporate a set of model performance indicators (MPIs) that describe the academic language proficiency expectations for ELL students at five different grade-level clusters and in five different content areas.

The grade level clusters of the standards used for test development purposes are Pre-K–K, 1,-2, 3–5, 6–8, and 9–12. Test forms follow this grade-level clustering as well, though the K test is appropriate only for kindergarten, and not for pre-kindergarten. The WIDA framework recognizes the continuum of language development within the four language domains across six proficiency levels.

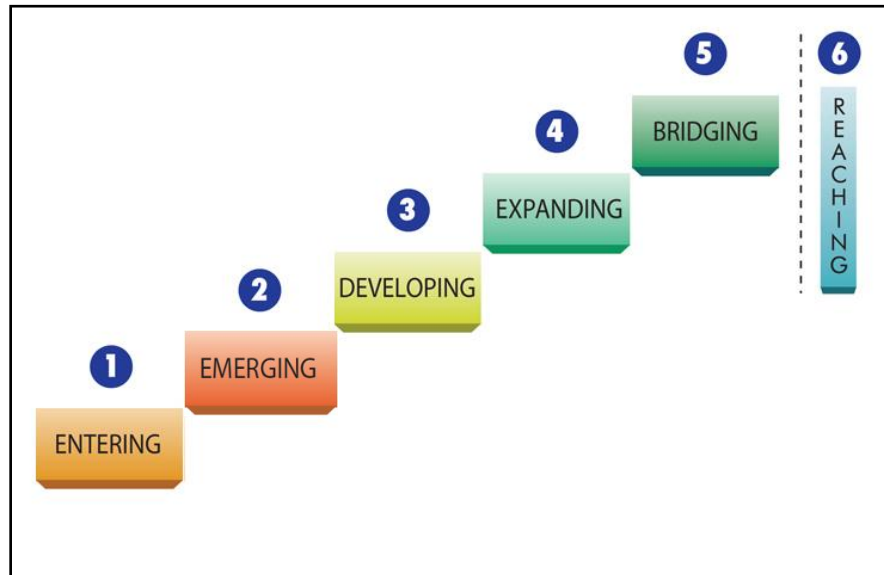
The six levels depicted in Figure 21 describe the spectrum of a learner's progression from knowing little to no English to acquiring the English skills necessary to be successful in an English-only mainstream classroom.

---

<sup>2</sup> Although ACCESS for ELLs includes a kindergarten test, this *Test and Item Design Plan* is concerned with the summative test and screener for **grades 1-12 only**. The development of a new kindergarten test is not funded under the ASSETS grant, nor is the development of the Alternate ACCESS for ELLs.

<sup>3</sup> In 2012, WIDA recharacterized its English Language Proficiency Standards as English Language Development Standards to emphasize that the process of language development is fluid and flexible.

Figure 11: The six proficiency levels



The ACCESS for ELLs test incorporates items assessing academic language from each of the five content areas of the standards. The first of these is Social and Instructional language (SIL), which incorporates proficiencies needed to deal with the general language of the classroom and the school. The other standards are content-based, and include the language of English Language Arts (LoLA), the language of Mathematics (LoMa), the language of Science (LoSc), and the language of Social Studies (LoSS).

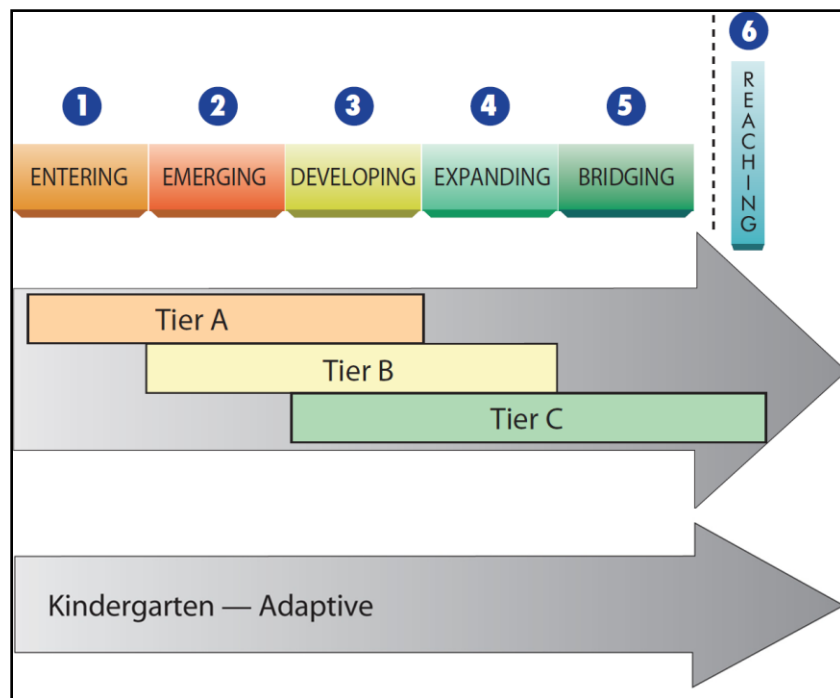
For each grade-level cluster, the standards specify one or more MPIs for each content area within each of the four language domains: Listening, Speaking, Reading, and Writing. Test items are written to specifications that address particular performance indicators. Each performance indicator and the items developed from it address one of five proficiency levels, which, in combination, describe a continuum of language proficiency development.

### A1.3 Tiers

The goal of the ACCESS for ELLs test is to allow students to demonstrate their level of proficiency through the MPIs. However, a test with questions assessing each and every MPI would be far too long to fit in any reasonable testing session. Too many easy questions make a test boring, while too many hard ones make it frustrating. It is important to avoid both extremes to achieve a balanced and reliable test. To ensure that an individual student takes the tests that is appropriate for his or her English language proficiency, test items are presented in 3 three tiers for each grade-level cluster; these are designated as Tiers A, B, and C. Figure 22 shows how the different tiers map to the proficiency levels.



Figure 12: Tiers and Proficiency Levels



As illustrated in Figure 22, the tiers overlap: Tier A covers levels 1-3, Tier B covers levels 2-4, and Tier C covers levels 3-5. The Tier C test also includes some items that are slightly more difficult than proficiency level 5. The kindergarten test is not tiered.

Each tier, of course, is only able to discriminate performance on its portion of the proficiency scale, so to make sure that the whole ACCESS for ELLs test works as intended, it is necessary to place each student into the tier that best matches his or her proficiency level. The decision as to where the student currently falls on the scale is best made by the student’s teachers, based on the information they have about the student’s language proficiency, including performance on other language tests, such as the WIDA-ACCESS Placement Test (W-APT™).

For students lacking other test scores, the WIDA Consortium provides a protocol—shown in Figure 23—for placing students into the appropriate tier of the test. Matching students to tiers appropriately will maximize accuracy and validity of the results. Note that English language learners must meet at least ONE of the criteria listed under each tier to qualify for that tier.

Figure 13: Tier Descriptions

Level 1 Entering	Level 2 Emerging	Level 3 Developing	Level 4 Expanding	Level 5 Bridging	Level 6 - Reaching
<b>TIER A</b> is most appropriate for English language learners who: <ul style="list-style-type: none"><li>• have arrived in the U.S. or entered school in the U.S. within this academic school year without previous instruction in English, OR</li><li>• currently receive literacy instruction <b>ONLY</b> in their native language, OR</li><li>• have recently tested at the lowest level of English language proficiency</li></ul>					
		<b>TIER B</b> is most appropriate for English language learners who: <ul style="list-style-type: none"><li>• have social language proficiency and some, but not extensive, academic language proficiency in English, OR</li><li>• have acquired some literacy in English though have not yet reached grade level literacy</li></ul>			
		<b>TIER C</b> is most appropriate for English language learners who: <ul style="list-style-type: none"><li>• are approaching grade level in literacy and academic language proficiency in the core content areas, OR</li><li>• will likely meet the state's exit criteria for support services by the end of the academic year</li></ul>			

## A1.4 Test administration times

The target administration times for each section of the test are:

- Listening: 25 minutes, machine scored
- Reading: 35 minutes, machine scored
- Writing: up to 1 hour, rater scored
- Speaking: up to 15 minutes, administrator scored

The group-administered tests at grade levels 1-12 are given in two 75-minute sessions: Session 1 for the Listening and Reading tests and Session 2 for the Writing test. The Speaking test must be administered to students individually in a separate third session.

The entire kindergarten test is administered individually, and, depending on the student's proficiency, may take up to 45 minutes to administer. More detailed instructions for administering ACCESS for ELLs are contained in the test administration manuals available on the WIDA website.